THE OPEN UNIVERSITY
A SCIENCE FOUNDATION COURSE

# UNIT 24 DNA: MOLECULAR ASPECTS OF GENETICS



THE OPEN UNIVERSITY PRESS

# THE SCIENCE FOUNDATION COURSE TEAM

Steve Best (Illustrator)
Geoff Brown (Earth Sciences)
Jim Burge (BBC)
Neil Chalmers (Biology)
Bob Cordell (Biology, General Editor)
Pauline Corfield (Assessment Group and Summer School Group)
Debbie Crouch (Designer)
Dee Edwards (Earth Sciences; S101 Evaluation)
Graham Farmelo (Chairman)
John Greenwood (Librarian)
Mike Gunton (BBC)
Charles Harding (Chemistry)
Robin Harding (Biology)
Nigel Harris (Earth Sciences, General Editor)
Linda Hodgkinson (Course Coordinator)
David Jackson (BBC)
David Johnson (Chemistry, General Editor)
Tony Jolly (BBC, Series Producer)
Ken Kirby (BBC)
Perry Morley (Editor)
Peter Morrod (Chemistry)
Pam Owen (Illustrator)
Rissa de la Paz (BBC)
Julia Powell (Editor)
David Roberts (Chemistry)
David Robinson (Biology)
Shelagh Ross (Physics, General Editor)
Dick Sharp (Editor)
Ted Smith (BBC)
Margaret Swithenby (Editor)
Nick Watson (BBC)
Dave Williams (Earth Sciences)
Geoff Yarwood (Earth Sciences)

Consultants: Keith Hodgkinson (Physics)
Judith Metcalfe (Biology)
Pat Murphy (Biology)
Irene Ridge (Biology)
Jonathan Silvertown (Biology)

External assessor: F. J. Vine FRS

Others whose S101 contribution has been of considerable value in the preparation of S102:

Stuart Freake (Physics)
Anna Furth (Biology)
Stephen Hurry (Biology)
Jane Nelson (Chemistry)
Mike Pentz (Chairman and General Editor, S101)
Milo Shott (Physics)
Russell Stannard (Physics)
Steve Swithenby (Physics)
Peggy Varley (Biology)
Kiki Warr (Chemistry)
Chris Wilson (Earth Sciences)

The illustration on the front cover is taken from *Cassell's Natural History*, Vol. IV (1883).

The Open University Press, Walton Hall, Milton Keynes, MK7 6AA.
First published 1988.



Designed by the Graphic Design Group of the Open University.

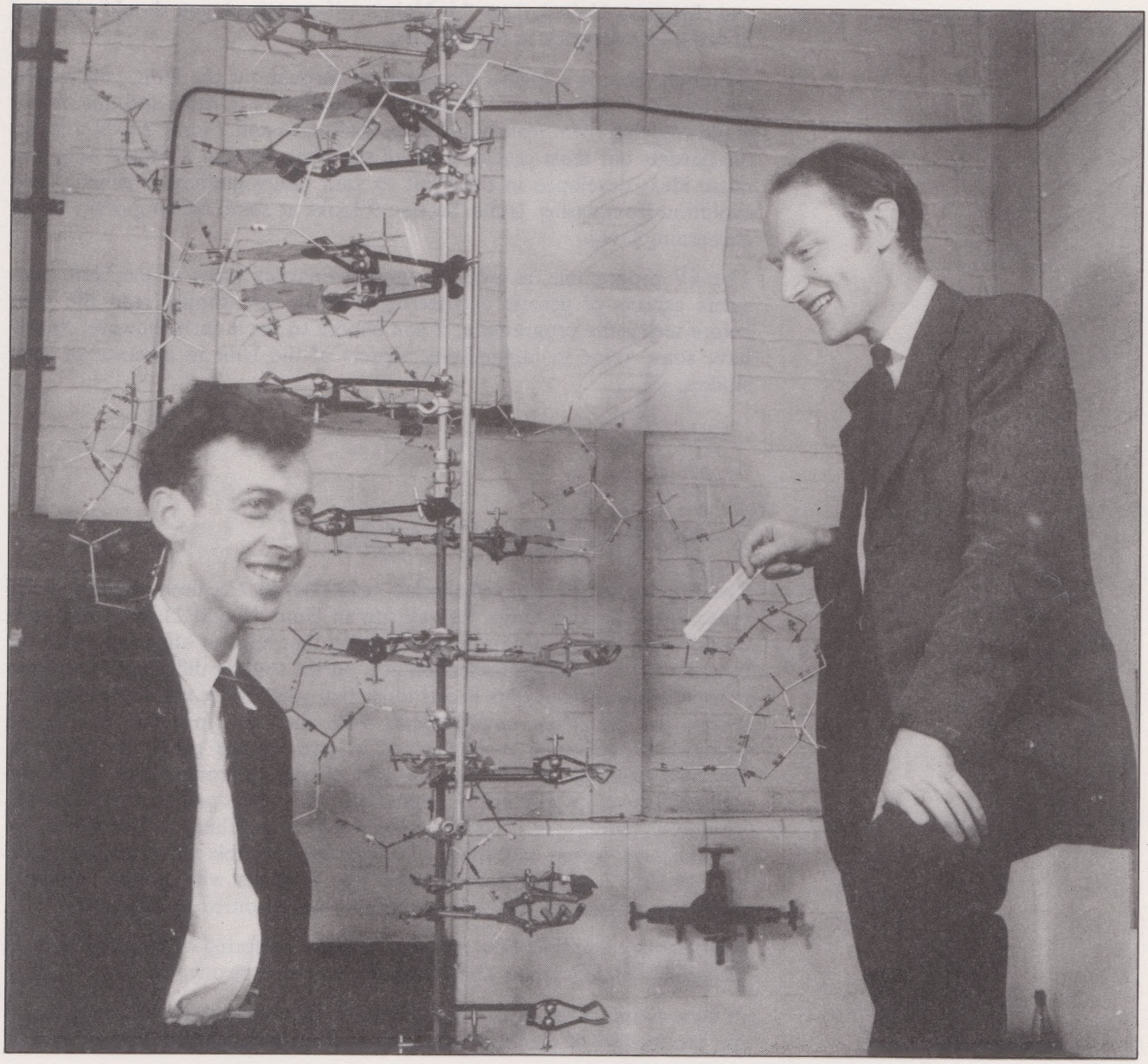
Filmset by Santype International Limited, Salisbury, Wiltshire, printed by Thomson Litho, East Kilbride, Scotland.

ISBN 0 335 1637 8

The text forms part of an Open University Course. For general availability of supporting material referred to in this text please write to: Open University Educational Enterprises Limited, 12 Cofferidge Close, Stony Stratford, Milton Keynes, MK11 1BY, Great Britain.

Further information on Open University Courses may be obtained from the Admissions Office, The Open University, P.O. Box 48, Walton Hall, Milton Keynes, MK7 6AB.

1.1



Watson and Crick with their original DNA model. Cambridge, 1953.

DEOXYRIBONUCLEIC ACID, DNA

# STUDY GUIDE

There are two components to this Unit—the text and the TV programme 'DNA'. The Unit is about the *molecular* biology that underlies genetics and evolution—how DNA, the heritable material in genes, is connected through RNA and the synthesis of proteins to the phenotype of an organism. The Unit builds on ideas about DNA that you have met in previous biology and chemistry Units but, as this Unit is self-contained, you should not need to revise or refer back to those Units.

Although there is a logical development throughout the Unit, you should find that each of Sections 1 to 5 is a complete story in itself, and this should make studying them reasonably easy. Sections 6 and 7 form a pair, and are the hardest but most crucial parts of the Unit. Section 8 should help you relate ideas developed in this Unit to your understanding of genetics and evolution from earlier Units. Section 9 looks at some contemporary issues concerning DNA.

The TV programme, as well as summarizing key areas of the Unit, considers aspects of genetic engineering. Ideally, you should read the Unit before seeing the programme; if you choose to see it early, however, it will have some value in highlighting aspects of the Unit in advance of your studies.

# 1 INTRODUCTION

The importance of **deoxyribonucleic acid (DNA)** has been stressed many times in the biology Units. As you saw in Units 19 and 21, evolution proceeds by the action of natural selection on the range of offspring produced by every species. Natural selection acts on the slightly differing phenotypes of each offspring, and thus selects those that can survive and reproduce best within the prevailing environment. You also saw in Unit 20 that an organism's phenotype is determined in part by the genes the organism inherited from its parents, and in part by the environment in which the organism develops and grows. Blond hair could be a genetic trait inherited from your parents, or it could be the result of bleaching by the Sun or by chemicals. However, only those traits that are heritable are significant as far as evolution is concerned; different sets of genes will produce different phenotypes which, in turn, are favoured or not favoured by selection.

Unit 20 also described how it is that different offspring of the same parents have different genotypes. These differences arise in two ways:

(i) as a consequence of the reshuffling (recombination) of existing genes in the production of gametes—the action of independent assortment and of crossing-over;

(ii) through new genes being formed as a result of mutation.

A key question is, 'How do the genes determine phenotype?' In Unit 22 you saw that phenotype is determined, in the main, by the nature of the proteins in the organism. It is the structure, and hence function, of the proteins that is determined by the genetic material—DNA. The idea that 'DNA makes RNA makes protein' is now very familiar to you. What, however, are the molecular processes underlying these relationships?

A number of other questions about genetics may have occurred to you. You know that gametes are produced by meiosis, and that each meiotic division involves (*en route*) the production of two joined chromatids from each chromosome. The formation of many somatic cells from one zygote depends on mitosis, and this too involves (*en route*) the doubling up of each chromosome to give two joined chromatids. Making chromatids ('two from one') requires that DNA, the genetic substance, be duplicated. This leads to another question: 'How is this done?'.

Perhaps you have also wondered about an apparent paradox related to mitosis. If mitosis copies the genetic message faithfully from each cell to its daughter cells, it would seem that the same total package of genetic information should be present in *all* cells derived from a particular zygote. DNA, we are told, determines phenotype, and the DNA content of all the cells in an organism is the same. Yet the phenotypes of muscle cells, blood cells, nerve cells, skin cells and so on are plainly different! How can this be?

Even more fundamentally, how are we so sure that DNA *is* the genetic substance of cells? Before the 1940s, scientists had some very different ideas about this, so what is the evidence that makes it now so certain that DNA has this crucial role in all organisms except a few viruses? And if DNA really is the genetic substance, what kind of structure does the DNA molecule possess that gives it the rather marvellous properties on which its functions depend—how does it replicate, how does it contain heritable information, and how does it change as a result of mutation? Thinking back to the Units on evolution, you will recall that both mutations and the inheritance of genes are essential parts of evolutionary theory. Thus the questions we have just asked are central to biology, and are the concern of Sections 2–8.

The reassuringly tidy story about DNA told in those Sections has been accepted, taught and learnt for many years. Gradually, however, more and more experimental work has led to new complexities and subtleties in what we know about the structure and functions of DNA, and, indeed, to the application of these developments to human use. So, in Section 9, you will see how recent research has caused us to reconsider the straightforward model of DNA structure and function presented in the earlier Sections.

The paragraphs above have posed a series of questions that are answered in the text. A list of these, expressed in shortened form, provides a useful guide through the Unit. Together with the Unit Sections in which they are discussed, these questions are:

1 What is the chemical nature of a gene? (*Sections 2 and 3*)
2 Why are cells different from each other? (*Section 4*)
3 How do genes replicate? (*Sections 3 and 5*)
4 How does genotype influence phenotype? (*Sections 6 and 7*)
5 How does mutation change genes? (*Section 8*)

Section 9, as noted above, looks at new ideas and applications, and Section 10 provides a brief summary of the Unit in which we find out whether the five questions above have been adequately answered.

# 2 DNA

So far we have made the assumption that genetic information *is* coded in the DNA molecule. As you will realize after reading this Unit, the experimental evidence for this is now so overwhelming that it hardly needs further thought. However, this reassuring state of affairs was not always the case. The experiments on bacterial transformation mentioned in the TV programme 'Darwin and Diversity' provided some of the first clues that DNA is the genetic substance. You will recall that chemicals from one strain of bacteria were able to 'transform' other strains of bacteria. The molecule involved in this transformation was found to be DNA. We will now briefly consider some of the other evidence that has been used to confirm the theory that DNA carries genetic information, and, in so doing, address Question 1 of the previous Section.

Viruses, as you know from Unit 19, have no independent life of their own. They are parasites of animal, plant and bacterial cells. They have a very simple structure—each virus is just a protein coat surrounding nucleic acid. Work with some DNA-containing viruses, namely those that parasitize bacteria, provided key evidence on the role of DNA as genetic material. One such virus is known as $T_2$.

## 2.1 EVIDENCE FROM VIRUSES

The $T_2$ virus (Figure 1a) can 'hijack' the components of a bacterial cell and make them manufacture more $T_2$ viruses. In Figure 1b, the outer tail section of the virus has contracted and the DNA inside the virus has been 'injected' into the bacterium. The presence of the viral DNA causes the *E. coli* cell to manufacture $T_2$ viruses.

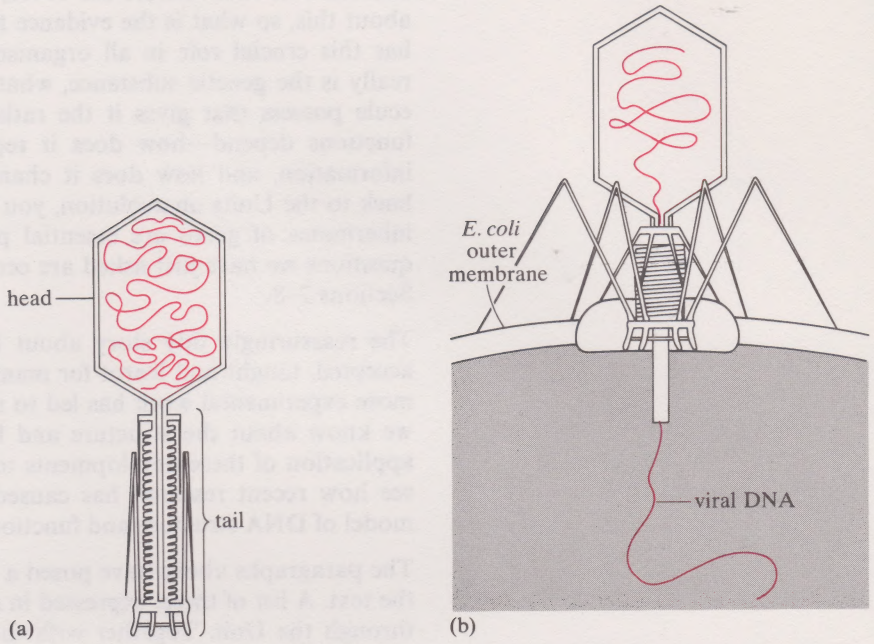



FIGURE 1 (a) Structure of the $T_2$ virus. (b) Process of injection of viral DNA into a bacterium.

A single virus can produce a hundred or more descendants in this way in as little as twenty minutes. Figure 2 shows the overall process of viral infection of a bacterium.

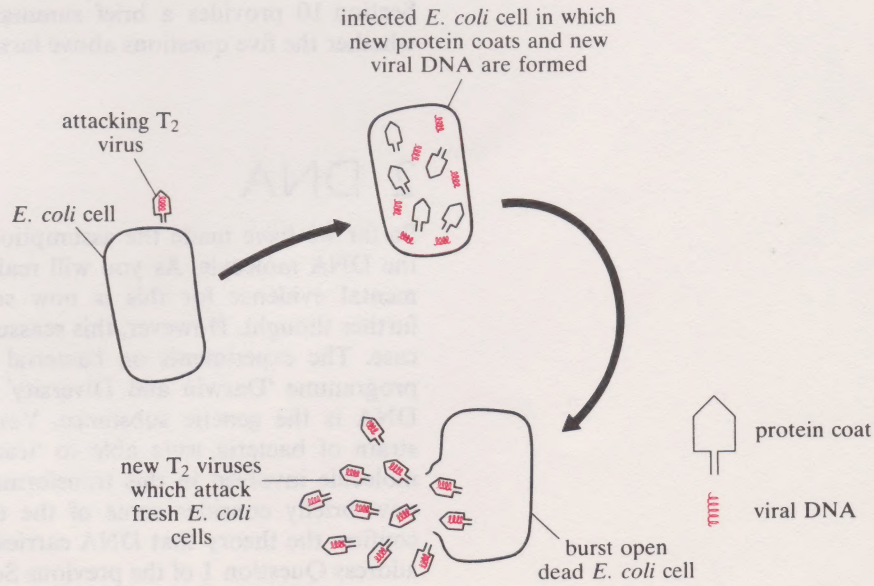



FIGURE 2 Infection of *E. coli* by $T_2$.

As a result of the virus infection, some of the enzymes in the bacterium somehow cease their normal job and instead begin to make new viral protein and new viral DNA. That is, the normal genetic message in the bacterial cell has been changed by the $T_2$ genetic message.

The question is, *which* part of the virus provides the genetic message? Is it the whole virus, the protein coat, or just the DNA? Alfred Hershey and Martha Chase, working in America, discovered the answer in 1952 by some very simple experiments. They found that when the virus infected a bacterium, the protein coat was left outside and the DNA entered the bacterium. This indicated strongly that the *DNA*, and not the protein coat, carried the viral genetic information.

## 2.2 DNA: THE GENETIC SUBSTANCE OF ORGANISMS

Look at Figure 3. You can see that the somatic cells (Unit 20) have the same number of *pairs* of genes as in the zygote. Gametes, produced by meiotic divisions of some particular somatic cells in adult organisms, have the same number of *single* genes—one from each pair of genes in the somatic cells.

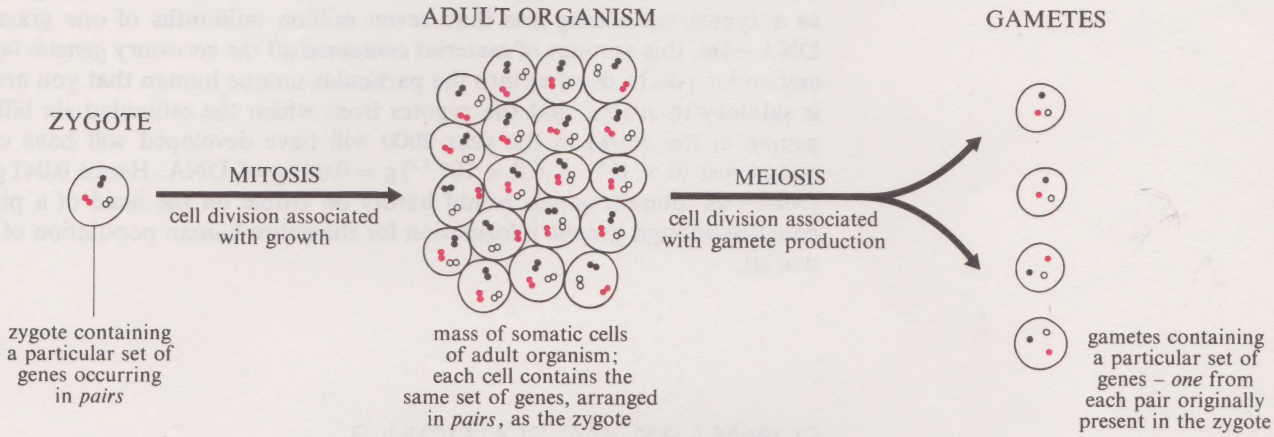



FIGURE 3 Genes and cell division: from zygote to gametes.

If DNA is the substance that carries the genetic information, what prediction would you make about:

(a) the mass of DNA in every somatic cell of a given organism, before the chromosomes are duplicated to chromatids;

(b) the mass of DNA in a zygote of that organism; and

(c) the mass of DNA in each of the gametes of that organism?

As Figure 3 shows, every somatic cell of a given organism contains the same set of pairs of genes. So, if the genetic substance is DNA, then one would predict that the mass of DNA in every nucleus of every somatic cell would be the same, in a given organism. Similarly, as the zygote has the same complement of gene pairs, one would predict that the mass of DNA in the zygote nucleus would be the same as that in the nucleus of each somatic cell derived from that zygote. The mass of DNA in a cell just prior to mitosis (when two chromatids have formed from each chromosome) will have twice the amount of DNA as in the non-dividing somatic cell. However, the gametes (having half the total number of genes, one from each pair) should have half the mass of DNA per nucleus, compared to that in a non-dividing cell. Table 1 shows how far these predictions are confirmed by determination of the mass of DNA in the nucleus for a number of different non-dividing cells.

TABLE 1 Average DNA content of a nucleus, for various organisms

| Organism | Red blood cell*/ picograms† per nucleus | Liver cell/ picograms per nucleus | Gamete (sperm)/ picograms per nucleus |
|---|---|---|---|
| domestic fowl | 2.3 | 2.4 | 1.3 |
| carp | 3.5 | 3.4 | 1.7 |
| toad | 7.3 | 7.4 | 3.7 |

* Unlike humans, the red blood cells of these species have nuclei.
† 1 picogram (pg) is $10^{-12}$ g.

You can see that, allowing for experimental uncertainty, these actual values fit in with our predictions. Estimates of the DNA content of nuclei from a wide range of species show a similar pattern. In all species studied, the DNA content of their somatic cells is virtually identical—and twice that found in the nuclei of the gametes of those species. For instance, the mass of DNA in a human zygote (6.8 picograms) is twice that in a human gamete (3.4 picograms). Some studies give a slightly different value for the zygote or somatic cells, but the gamete always has almost exactly half the mass of DNA compared with the somatic cells.

The incredible smallness of these masses deserves comment. You began life as a zygote containing less than seven million millionths of one gram of DNA—yet, this amount of material *contained all the necessary genetic information* for you to develop into the particular unique human that you are. It is salutary to realize that the zygotes from which the estimated six billion people in the world in the year 2000 will have developed will have contained just $(6 \times 10^9) \times (6.8 \times 10^{-12})\,g = 0.041\,g$ of DNA. Hence 0.041 g of DNA—an amount which would hardly be visible on the head of a pin—contains enough genetic information for the entire human population of the world!

## SUMMARY OF SECTION 2

1 DNA contains the genetic information of most living organisms. (The only exceptions are some viruses in which a different nucleic acid, RNA, is used as genetic material.)

2 Work with viruses confirmed that, during viral infection of bacteria, it is the nucleic acid (not the protein coat) that carries the viral genetic information.

3 Analyses of the amount of DNA in somatic cells as compared with that in the gametes show that, within the limits of experimental uncertainty, the mass of DNA in each gamete is half that in each of the somatic cells. This correlates with the fact that gametes have half the number of chromosomes compared with somatic cells.

**SAQ 1** If protein and not DNA were the genetic substance of the virus used in Hershey and Chase's experiment, predict what the observations in their experiment would have been.

**SAQ 2** (a) If trout sperm contains 2.7 picograms of DNA per nucleus, what mass of DNA would you expect to find in a non-dividing trout kidney cell?

(b) If the non-dividing cells lining the intestine of a turtle each contain 5.3 picograms of DNA per nucleus, what mass would you expect to find in

(i) an unfertilized turtle egg;

(ii) a single turtle sperm; and

(iii) a group of 32 cells formed from a turtle zygote?

# 3 THE STRUCTURE OF DNA

Though it is now plain that DNA is the genetic substance, we clearly need to explore its structure—its 'chemical nature', in the words of Question 1—in some detail. This Section will allow you to see that, despite the apparent complexities, DNA is a very simple system which both allows for the accurate replication of the DNA and provides the basis for a coding system that can determine the structure of other biological molecules. That is to say, we can start to see how DNA—the genotype of an organism—can influence the phenotype of that organism.

## 3.1 THE RELATIONSHIP BETWEEN THE STRUCTURE AND FUNCTION OF DNA

What is an organism's phenotype? Earlier Units have provided you with some kind of answer. It is the sum of an organism's characters, in terms of overall structure and biochemical, physiological and behavioural function. All these aspects of phenotype depend on the organism's chemical composition and the kind of biochemical reactions that go on inside it. It is this link between the chemistry of the organism and its structure and functions that is the key to how evolution can occur. For instance, the different coloration of the forms of the peppered moth (Unit 19) is ultimately dependent on chemical processes that determine the overall phenotype. These different colour forms will have slightly different biochemical processes, causing different coloration. The different wing colours have arisen at various times by mutation.

From Unit 22 it is clear that chemical structure and function, and hence phenotype, depend on proteins. All enzymes are proteins, and you will remember that specific enzymes bring about specific reactions that affect specific degradations or biosyntheses. However, besides enzymes there are other proteins of great phenotypic importance (e.g. haemoglobin, muscle protein, and protein hormones such as insulin). As you know, all these proteins are functionally different because of their different structures. Ultimately the structural differences depend solely on each protein's unique primary structure. This primary structure is the specific sequence of particular amino acids along the length of the protein molecule.

We know that DNA carries the genetic message that must be copied when new cells are made, that DNA must work through protein structure, and that the structure of DNA must be able to vary if evolution is to occur. From these known facts, we can list the points that *must* be characteristics of DNA structure. These are:

(i) DNA must be able to replicate; that is, produce perfect copies of itself.

(ii) DNA must contain 'instructions' within it for assembling amino acids in precise sequences.

(iii) DNA must be capable of alteration through mutation.

With these points in mind, let us now look at the molecular structure of DNA.

## 3.2 THE MOLECULAR STRUCTURE OF DNA

In this Section, a description of the double-stranded helical nature of DNA is built up by the series of diagrams in Figures 4–8 and Figure 10. You will find that you need to refer to these often in subsequent parts of the Unit, and you may therefore find it worthwhile to memorize the key features (though you will not be expected to remember them in the exam). In contrast, you should *not* memorize the structures in Figures 9 and 11!

DOUBLE HELIX

DEOXYRIBONUCLEOTIDE

NUCLEOTIDE

POLYDEOXYRIBONUCLEOTIDE

DEOXYRIBOSE

BASE

ADENINE, A

GUANINE, G

CYTOSINE, C

THYMINE, T

BASE-PAIRING RULES

COMPLEMENTARY BASES

PURINE BASES

PYRIMIDINE BASES

HYDROGEN BONDS

The main details of the structure of DNA were described by James Watson and Francis Crick in 1953. Their work was an epoch-making feat in biological terms, and in 1962 it gained for them and for Maurice Wilkins (a contributor to the task) the Nobel Prize for medicine and physiology. They found that DNA is a **double helix** consisting of two polymeric strands wound around each other. Although in precise terms the analogy is inexact, you can usefully consider the DNA double helix to be like a length of cheap electric flex (Figure 4).

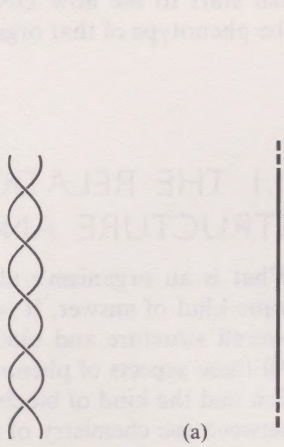


FIGURE 4 A simplified representation of the DNA double helix.

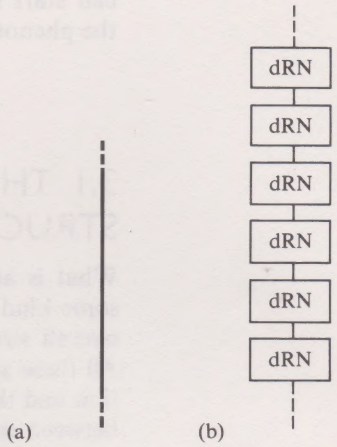
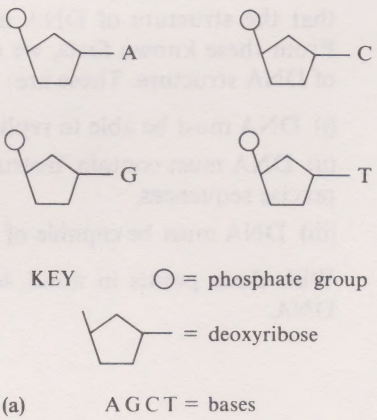



FIGURE 5 (a) A straightened-out single strand from the double helix shown in Figure 4.
(b) Polydeoxyribonucleotide structure of a single strand of DNA.

First let us consider the structure of each separate strand, and then see how the two fit together. Each DNA strand is a polymer of **deoxyribonucleotides** (usually abbreviated to **nucleotides**). Indeed, an alternative name for a DNA molecule is a **polydeoxyribonucleotide**. So, instead of representing the single strand by a line as shown in Figure 5a, we can draw the fuller structure shown in Figure 5b. Each box here represents one deoxyribonucleotide.

As you know from Units 17–18, a deoxyribonucleotide itself consists of three covalently linked parts: a sugar (**deoxyribose**), a phosphate group, and a nitrogen-containing ring structure called a **base**. As there are four different kinds of base in DNA—**adenine (A)**, **guanine (G)**, **cytosine (C)**, and **thymine (T)**—it follows that there are four different kinds of deoxyribonucleotide. These are shown in a simplified form in Figure 6.



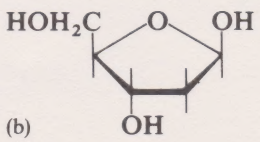




FIGURE 6 (a) Outline structures of the four deoxyribonucleotides. (b) The full structure of deoxyribose.

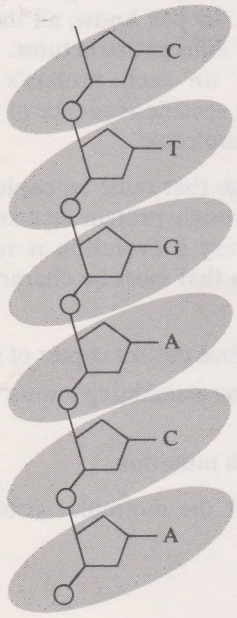



FIGURE 7 The polydeoxyribonucleotide structure of DNA shown in more detail than in Figure 5b. Each shaded oval is one deoxyribonucleotide. Note that the backbone of the strand consists of alternate phosphate and deoxyribose components (see key to Figure 6). The sequence of bases here is simply illustrative: we could have chosen any sequence.

We can use the information in Figure 6 to draw the structure of a DNA strand in Figure 5b in the slightly fuller way shown in Figure 7. It is the particular sequence of bases along the strand that makes one molecule of DNA different from another and provides the way in which the DNA molecule contains genetic information; more of that later.

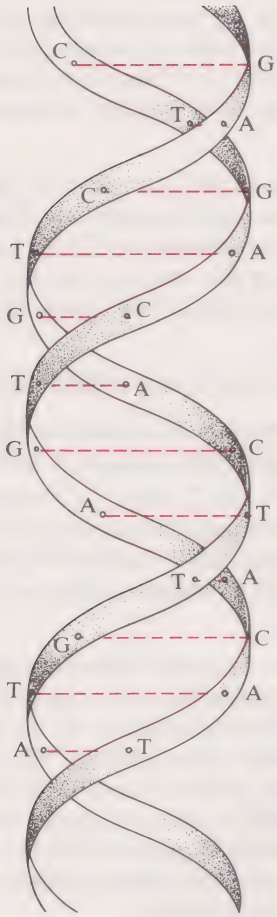



FIGURE 8 A schematic diagram of a DNA double helix, showing spiralling backbones of sugar–phosphate, and paired bases. The bonds between adjacent nucleotides in each backbone are strong covalent bonds, shown here as broad ribbons. The bonds between the bases (shown in red) are weak hydrogen bonds; we shall return to these later.

What about the other strand? This is exactly the same in terms of its phosphate–deoxyribose backbone, although there is something very special about the sequence of bases along the second strand. This feature is of great importance in the way the two strands fit together into a double helix, and is of *crucial* importance in the biology of DNA replication.

It happens that, in the complicated three-dimensional structure of the double helix:

(a) the sugar–phosphate backbones form the outside of the double helix, and

(b) the bases, attached horizontally to each deoxyribose molecule, lie inside the double helix—rather like the steps of a spiral staircase.

However, there are strict stereochemical limits to the bases that may lie opposite each other in the double helix. The rules are that A must lie opposite (be paired with) T, and G must pair with C. From these **base-pairing rules**, you can see that A and T constitute a pair of **complementary bases**; similarly, C and G constitute another pair. A is said to be complementary to T (and T with A): in like manner, C is complementary to G and vice versa.

So, with these relationships in mind, we can represent the structure of the double helix in the way shown in Figure 8. You can see now that the simple flex analogy in Figures 4 and 5a is indeed an over-simplification!

Although the chemistry behind these pairing relationships is quite complicated, the base-pairing rules in fact depend on just two factors. The first factor is the nature (and hence the size) of the nitrogen-containing rings of four bases A, G, C and T. Each of the bases adenine and guanine (A and G) has two nitrogen-containing rings joined together into a relatively large double-ringed structure known as a purine ring; these bases are therefore called **purine bases**. Conversely, cytosine and thymine both have just one nitrogen-containing ring, and their smaller molecules are known as **pyrimidine bases**. All four of these bases are shown in Figure 9.

If we represent the larger purine rings by a large rectangle and the smaller pyrimidine rings by a smaller rectangle, you can see that, within the confines of a regular helical structure that is not distorted by bulges or constrictions, it will only be possible to have a purine ring opposite a pyrimidine ring, and vice versa, as shown in Figure 10. Putting two large rectangles or two small rectangles opposite each other would distort the backbone of the DNA molecule. It follows that A must pair with C or T, and G must pair with C or T; A–G pairing and C–T pairing are ruled out.

The second factor on which the base-pairing rules depend is the nature of the partially charged groups around each of the bases. The weak bonds linking the bases opposite each other in the double helix are familiar to you from Units 17–18.

□ Can you recall the name and nature of these bonds?

■ They are **hydrogen bonds**. A hydrogen bond arises between a covalently bound hydrogen atom bearing a partial positive charge and some other covalently bound atom bearing a partial negative charge. These 'other' atoms are electronegative ones such as O or N.

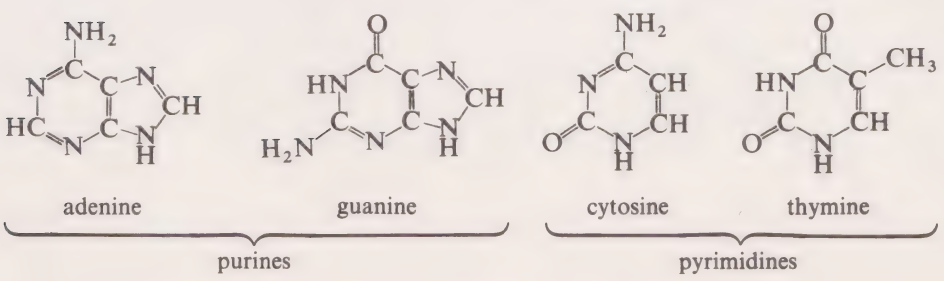



FIGURE 9 The ring structures of the four bases in DNA.

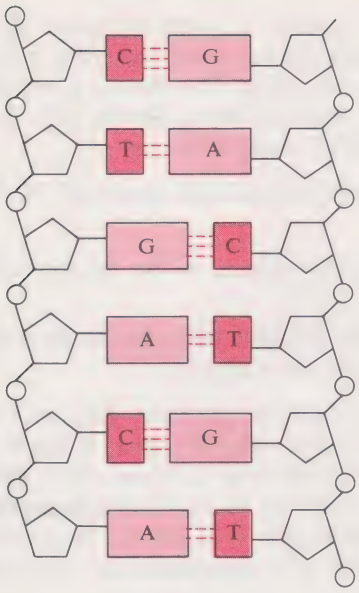



FIGURE 10 Purine and pyrimidine base pairing in a DNA double helix.

The structure of adenine is such that it is complementary to that of thymine and these bases pair with each other in a unique way, stabilized by *two* hydrogen bonds between them. Similarly, the structures of G and C are complementary and their pairing is stabilized by *three* hydrogen bonds between them. Any other pairing would permit fewer hydrogen bonds to form and the structure would be much less stable—but you do not need to dwell upon this point. The nature of A–T and G–C pairings is shown in Figure 11.

In ways which would have been unimaginable in 1953, we are now able to use computers to generate 3-D pictures of biological molecules. Plates 1 and 2 show some recent pictures of the molecular structure of DNA. Generating representations of molecular structures in this way is a far cry from the painstaking model building used by Watson and Crick—look back at the frontispiece to this Unit!

The main points you should remember about the structure of DNA are:

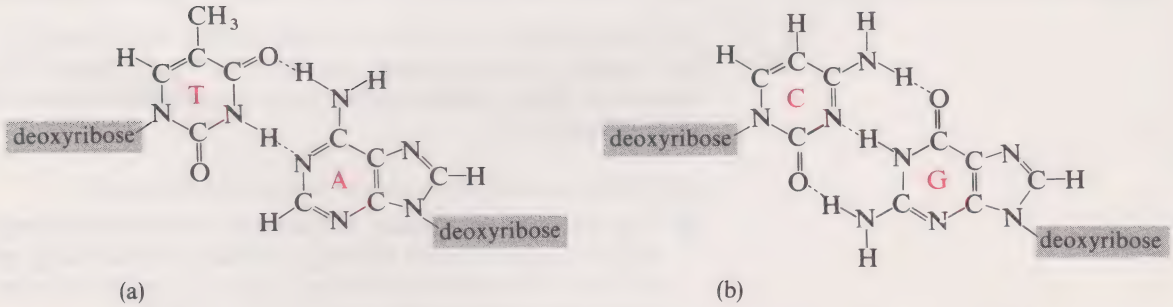
1 Purine–purine or pyrimidine–pyrimidine pairing is impossible because of the sizes of their rings. Only purine–pyrimidine ring pairing is possible.

2 The precise molecular structures of A, G, T and C make A–T and G–C the only possible combinations; two hydrogen bonds stabilize the A–T pair and three hydrogen bonds stabilize the G–C pair. Look back at Figure 10; you will see that the dashed lines, three or two in number, represent these (rather weak) hydrogen bonds.

The point has already been made that a sequence of four different kinds of base along a strand could perhaps provide information about the sequence of amino acids in a protein. There is another important feature of the structure of DNA that may not have escaped your attention: the way each strand of the double helix is complementary to the other, in the base-pairing sense, could be the basis of the way DNA is able to replicate itself. If this did occur to you, you will have re-created the thoughts of Watson and Crick when, in 1953, they published in the scientific journal *Nature* their famous paper containing this magnificent understatement:

> We wish to suggest a structure for the salt of deoxyribose nucleic acid (DNA). . . . It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

The relationship between the structure of DNA and its method of replication is the theme of Section 5.




(a) (b)

FIGURE 11 Base pairing by hydrogen bonding between: (a) adenine (A) and thymine (T); and (b) cytosine (C) and guanine (G). Hydrogen bonds are shown by dashed lines.

## SUMMARY OF SECTION 3

1 DNA consists of two polymeric strands wound round each other. These strands are composed of an outer backbone of phosphate and sugar molecules, with nucleotide bases projecting inwards from this backbone.

2 The sequence of bases provides a unique sequence within the DNA molecule which carries genetic information.

3 There are precise chemical rules that determine how the bases pair up between the two strands of DNA. Adenine pairs with thymine and cytosine pairs with guanine (A–T and C–G) to form complementary base pairs.

4 These pairings are determined by:

(a) the size and shape of the base—purines have a double ring structure and are larger than pyrimidines which have a single ring structure;

(b) the number of hydrogen bonds that can stabilize the structure—there are two hydrogen bonds between adenine and thymine, and three hydrogen bonds between cytosine and guanine.

**SAQ 3** (a) Fragment Z is part of a double helical molecule of DNA, and it is represented in simplified form (i.e. drawn straight rather than as a helix) in Figure 12. In fragment Z, what is:

(i) the ratio of adenine bases to thymine bases,

(ii) the ratio of guanine bases to cytosine bases, and

(iii) the ratio of purine (pu) bases to pyrimidine (py) bases?

| A | G | G | T | T | A | C | C | G | C | A | A | G |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T | C | C | A | A | T | G | G | C | G | T | T | C |

FIGURE 12 Fragment Z, which is part of a double helical molecule of DNA.

(b) If you were to consider *another* part of the same double helical DNA molecule, what predictions could you make about the A:T, G:C and pu:py ratios in that part?

**SAQ 4** In fragment Y of a DNA double helix there are found to be 80 purine bases, 23 cytosine bases and $n$ other bases. Calculate the total number of each of the following items in fragment Y:

(a) bases (of all types)

(b) thymine bases

(c) adenine bases

(d) guanine bases

(e) hydrogen bonds

(f) phosphate groups

(g) nucleotides

(h) deoxyribose molecules

(i) complementary base pairs

DIFFERENTIATION

NUCLEAR TRANSFER EXPERIMENTS

# 4 CELL GROWTH AND DIFFERENTIATION

From Unit 20 you know of the events that occur during mitotic cell division. An essential feature of mitosis is that, during its early stages, the number of chromosomes in the cell is effectively doubled; this means that when division occurs, each new cell will have the correct number of chromosomes. Thus the DNA must replicate, at some point, before each and every mitotic division. As a consequence, every cell in an organism is provided with an identical set of genes, and hence an identical set of DNA molecules. Despite this, there is a wide variety of cell types in most organisms. As we asked in Question 2 of the Introduction, how can this be? To put the answer to this question in perspective, we begin by reviewing the role of mitosis in the growth of organisms—the cell cycle.

## 4.1 THE CELL CYCLE

Most cells go through a continuous cyclic process of growth and division. A simple example is yeast cells growing, and increasing in number by mitotic cell division, in a weak solution of sugar. This will be familiar if you make wine at home. The yeast cells continue to divide and grow until the toxic products of metabolism inhibit growth. There is a build-up of ethanol in the solution, which eventually kills the yeast cells.

If we were to look at the process of DNA replication and cell division during one cell cycle of typical cells in most animals and plants, we would see a picture similar to that shown in Figure 13. This Figure shows that the time from the end of one mitotic division to the beginning of the next is about 24 hours. During this period the DNA content changes as shown. This diagram has four distinct phases.

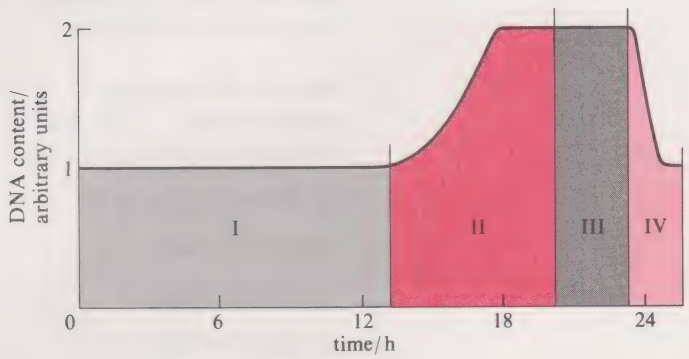



FIGURE 13 The phases of the cell cycle in a typical cell.

☐ Which phase is associated with DNA replication?

■ Phase II—it is during this phase that the DNA content of the cell increases by a factor of 2. Towards the end of this phase you would see the paired chromatids in the dividing cell. These are a consequence of DNA replication.

☐ Which phase is associated with cell division (mitosis)?

■ Phase IV—here there is a sudden *decrease* by a factor of 2 of the DNA within each cell.

☐ What is occurring in Phase I?

■ This is a period of cell growth prior to DNA replication.

Phase III is a period when reorganization of cellular components (such as the membranes within the cell) occurs, immediately before mitosis.

Cells which are constantly dividing must be continually replicating their DNA, and thus the cell cycle is a feature of the vast majority of cells. Figure 14, showing the cell cycle over four generations, brings us back to a paradox about mitosis—if DNA is fully replicated in the cell cycle, how is it that cells in the same organism are so different?

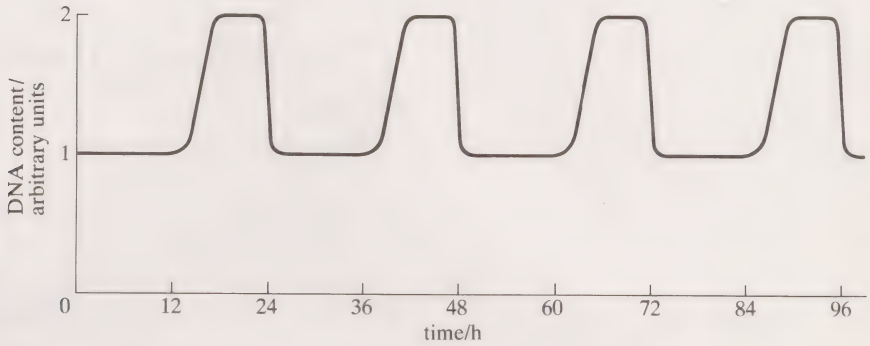



FIGURE 14 The cell cycle in a typical cell, over four generations.

## 4.2 GROWTH AND DEVELOPMENT

Growth and development are features of all organisms. We can appreciate that organisms grow from the small single cell after fertilization (the zygote) to the vast number of cells in the adult organism. We know that these cells in the adult organisms are also continually dying and being replaced by new cells. We also know that the form of an adult organism is dramatically different from that of the original zygote. Development implies changes in both shape and function for the organism.

All sexually reproducing organisms began life as a zygote. The DNA in the human zygote that you once were, carried all the information that—in conjunction with the environmental forces in your life so far—has made you the unique individual you are. Now, as well as the millions of mitotic cell divisions that have gone on—each of which was preceded by the perfect replication of that DNA—something else has happened to these cells.

As cell division occurs, certain cells become specialized. As a result, nerve cells, heart cells, liver cells, bone cells, blood cells and so on appear. The formation of cells different from each other and different from the original zygote is termed **differentiation**.

Given that all the cells in an organism have the same genetic material, how do these differences between the cells arise? You might deduce from what you know about cell division and the cell cycle that all cells would have the same genetic information; but is there any direct evidence to support this idea? It is possible to formulate a hypothesis that differentiation is due to the loss of all the DNA *except* that needed for the specific functions of the differentiated cell.

**Nuclear transfer experiments** carried out by the biologist John Gurdon in the 1960s provided good evidence that all cells do indeed carry the same genetic information and that there is no physical loss of material as a result of numerous cell divisions. From tadpoles of the toad (*Xenopus*), he took the nuclei from mature, differentiated cells lining the gut and inserted them into toad eggs from which the original nuclei had been effectively removed by irradiation (Figure 15). Although about 90% of the eggs with implanted tadpole gut nuclei remained unchanged until they eventually died, some began to develop in the manner of a normal toad zygote. Amazingly, bearing in mind that the egg contained the nucleus of an adult cell and not that of a zygote, a small proportion of the treated eggs developed into adult fertile toads. This showed, of course, that the nuclei of the fully differentiated tadpole gut cells retain all the genetic information present in the nucleus of the zygote from which they developed.

## CONTROL GENES

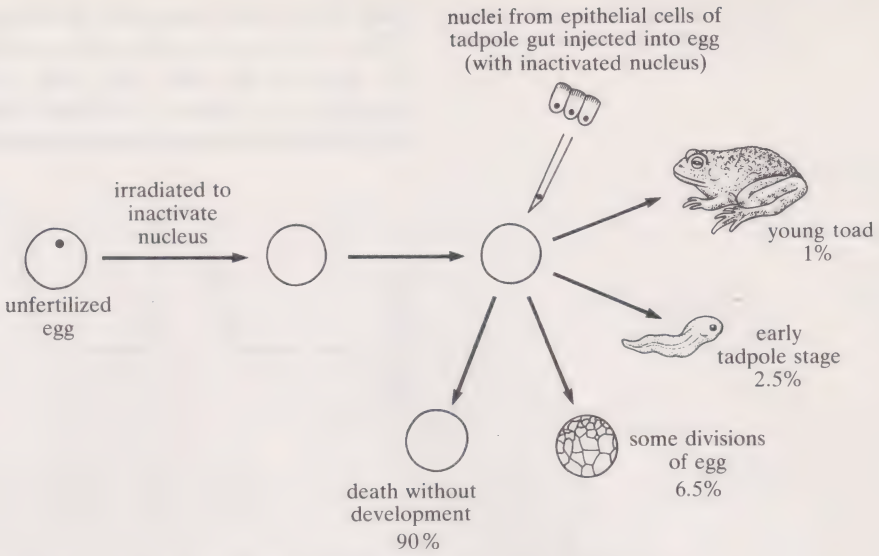



FIGURE 15 Outline of Gurdon's nuclear transfer experiments.

If we assume, from the results of experiments like these, that all the cells of an organism contain the same genes, we have to ask how cells in the same organism can be so different in structure and function. To answer this question, even to the limited extent of current knowledge, would require several Units! As an example of the complicated mechanisms which may be involved in the control of gene functions, we shall consider, very briefly and simply, how genes might be 'switched on and off'.

During the early life of all organisms, growth, development and cellular differentiation must be highly ordered and controlled: events must occur in the correct sequence and at the correct time. These sequential events result from control being exercised over the DNA in the cell. Small sequences of DNA are known to have a controlling function over the activity of the genes that make proteins. These areas of the DNA are known as **control genes**. These particular genes 'switch' other genes on or off, thereby giving rise to the chronological sequence of events that differs from tissue to tissue. Thus, for example, muscle cells make muscle proteins because those genes are switched on in these cells. The genes for muscle protein production in all other cells are switched off.

Although this idea of switching genes on or off gives a possible mechanism for differentiation, we are no further forward in explaining what it is that does the switching. Put another way, what is it that controls the control genes? It is known that the chemical environment of the cell and its position relative to other cells influence its development. In other words, the development of a cell is not solely a function of that cell's genetic make-up, but can be strongly influenced by its local environment. Cells in the region of the embryonic eye develop the structure, composition and properties of eye retinal cells, by switching on particular enzymes that catalyse certain biochemical reactions. In contrast, cells located in parts of the embryo destined to become arms and legs develop the typical fibres and contractile proteins of muscle; in these cells, chemicals characteristic of retinal cells are notable for their absence!

Current work on how the cell cycle is controlled, and on cancer cells (cells that grow with little differentiation), should give further insight into the control processes which coordinate growth and development. At present, though, it has to be said that the process of growth and development in complex multicellular organisms is poorly understood. In the not too distant future, however, it is likely that our knowledge of the detailed workings of genetic material at the molecular level will lead to a fuller understanding of the questions raised in this Section.

## SUMMARY OF SECTION 4

1 Cells that go through a series of divisions exhibit a cyclical pattern of events known as a cell cycle. The DNA content of cells during each cycle doubles just before cell division. DNA replication—which doubles the DNA content of the cell—is an essential prerequisite for cell division.

2 The development and differentiation of cells involves the ordered expression of parts of the genetic information within those cells.

3 There is no physical loss of genetic information as differentiation occurs.

4 The control of differentiation is likely to involve the ordered switching on and switching off of genes, by means of chemical signals.

**SAQ 5** Draw a simple diagram to show the relative changes in the DNA content of a cell during a period that starts with cell growth and ends with cell division. Indicate on your diagram the four phases of this cell cycle and outline the events which occur in each of these phases.

**SAQ 6** Which of the following statements is false?

(a) If no adult toads had developed as a result of Gurdon's nuclei transfer experiments, this would have indicated that differentiation must occur as a result of the loss of certain genes.

(b) If the nucleus of a phoenix zygote contained $x$ picograms of DNA, phoenix muscle cells $0.8x$ picograms of DNA per nucleus, the cells of phoenix liver $0.6x$ picograms of DNA per nucleus, and the nuclei of phoenix kidney cells $0.5x$ picograms of DNA, these data would strongly suggest that differentiation in the phoenix used to occur as a result of differential loss of genes.

(c) It is reasonable to suppose that cells within the human big toe have genes that code for the production of enzymes responsible for the biosynthesis of rhodopsin (the visual pigment found in the retina of the eye).

# 5 DNA REPLICATION

In the Introduction, Question 3 asked 'How do genes replicate?'. The first hint of an answer, discussed in Section 3, was a purely theoretical one. As you saw earlier, Watson and Crick were sure that a copying mechanism was inherent in the base-pairing characteristics of their newly elucidated structure for DNA. However, the first experimental evidence that each strand of the double helix acts as a template (i.e. rather like a mould) for the assembly of a complementary strand did not become available until 1958, five years after Watson and Crick's paper was published in 1953.

## 5.1 A THEORETICAL SCHEME FOR DNA REPLICATION

By 1953, the following was known about DNA:

1 DNA is a double helix in which each strand is composed of many nucleotides joined together.

2 Base A of the nucleotide in one strand must pair with base T of the opposite nucleotide in the other strand; and similarly for G and C.

3 The bonds linking paired bases are hydrogen bonds; these weak bonds are much more easily broken than strong covalent bonds.

4 The liquid medium surrounding the DNA molecule contains free nucleotides of all four bases.

SEMI-CONSERVATIVE REPLICATION

DAUGHTER DOUBLE HELICES

MESELSON AND STAHL EXPERIMENT

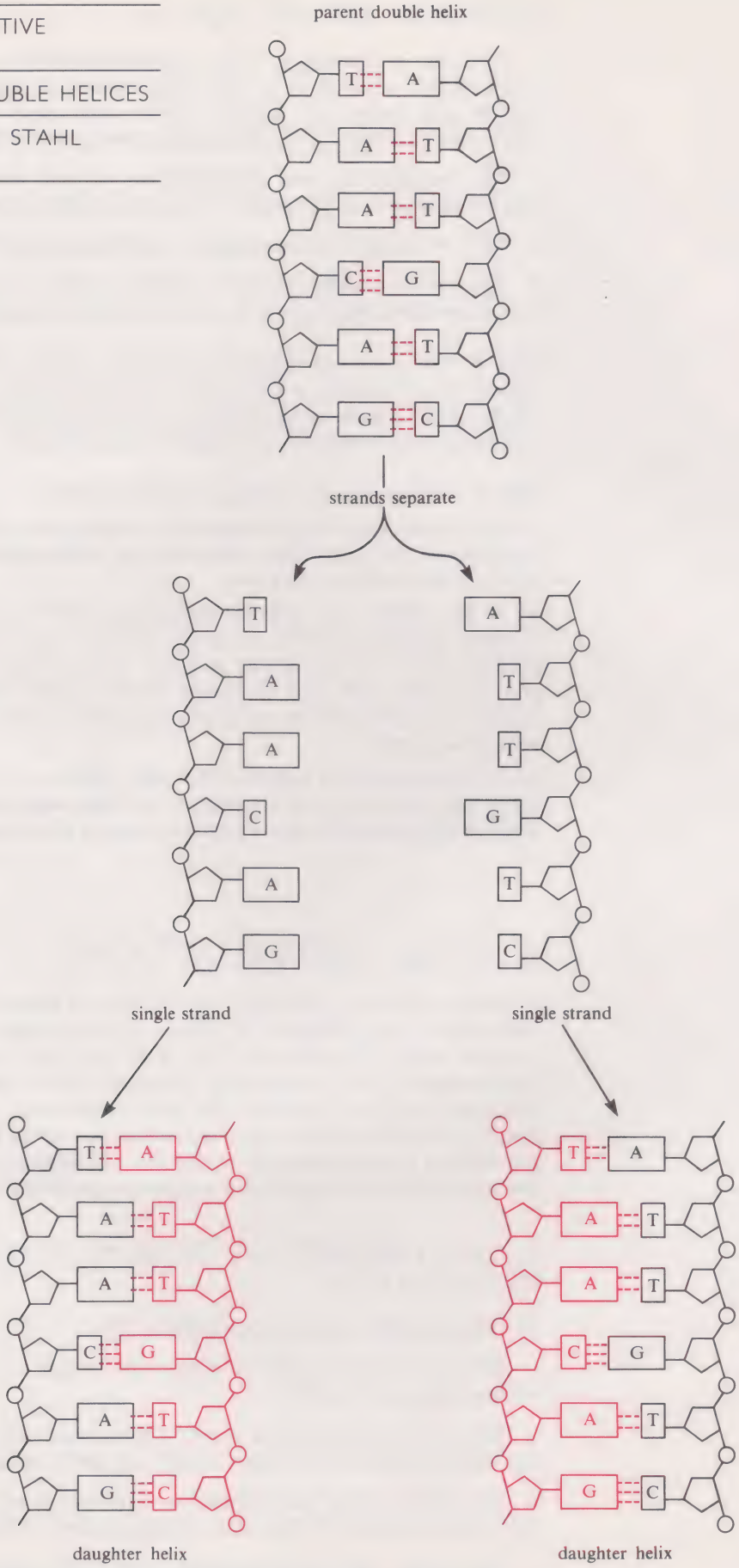



FIGURE 16 A theoretical scheme for DNA replication. Daughter helices and hydrogen bonds are shown in red.

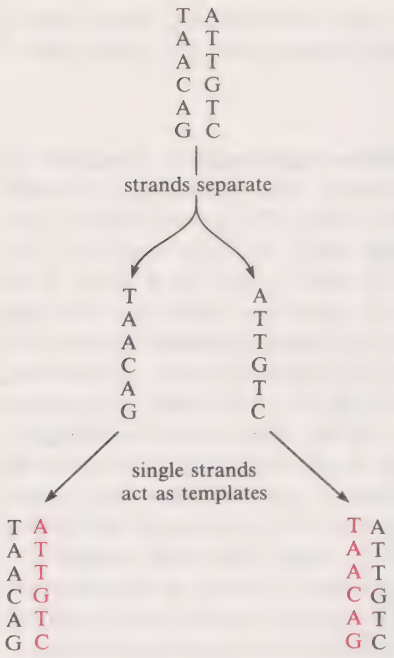



FIGURE 17 A simplified form of Figure 16. New strands of DNA are shown in red.

ITQ 1 Given all this information above, try and suggest a mechanism by which two new double helices, identical to each other, could be formed from one parental helix.

The way we think DNA replicates is summarized in Figure 16. If you blanch at the difficulty of drawing a diagram like that, then you will be relieved to see the conventionally acceptable 'short-hand' version shown in Figure 17. This Figure restates all the basic information given in Figure 16 but in a much simpler way, and it is much easier to write (and to remember!)

The scheme shown in Figures 16 and 17 is purely hypothetical—DNA *could*, in terms of its known structure, replicate in this way. Does this represent the truth and, if so, how do we know?

The scheme proposed above is known as **semi-conservative replication** because, according to this hypothesis, each of the two new DNA helices (called **daughter double helices**) in Figures 16 and 17 possesses one *old* strand and one *new* strand. Put another way, *half* of the old double helix is *conserved* in each of the new daughter double helices—hence the name 'semi-conservative'. What experimental evidence is there for this semi-conservative model?

## 5.2 EVIDENCE FOR SEMI-CONSERVATIVE REPLICATION

A classic experiment was conducted in 1958 by two American biologists, Matthew Meselson and Franklin W. Stahl, which provided strong evidence that DNA replication is of the semi-conservative type. **Meselson and Stahl** grew a culture of *E. coli* for many generations—bacteria divide about every 20 minutes—in a medium containing ammonium ions ($NH_4^+$) as a nitrogen source. But in these ions, the ordinary (light) nitrogen atoms $^{14}N$ had been replaced by the heavy isotope $^{15}N$, so giving $^{15}NH_4^+$ instead of $^{14}NH_4^+$. Living organisms metabolize the $^{14}N$ and $^{15}N$ isotopes at the same rates so, eventually, both strands of the DNA of *E. coli* became uniformly 'labelled' with the heavy $^{15}N$ isotope (Figure 18a). Because both strands were so labelled, it was described as 'HH DNA' instead of 'LL DNA' in the normal form. (Here, H stands for heavy and L for light.)

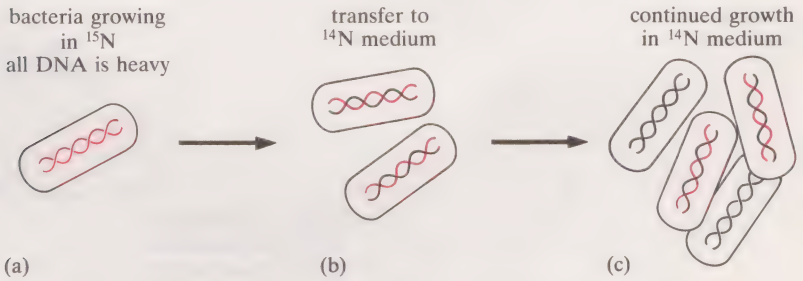



FIGURE 18 Change from 'heavy' to 'light' DNA, through two generations.

☐ Where are the N atoms in the DNA?

■ In the ring structures of the bases.

Some of these labelled *E. coli* cells (we'll call them the parent generation, generation 0) were then placed in a fresh medium containing unlabelled (light) ammonium ions as the source of nitrogen, and were allowed to divide just once (Figure 18b). Thus, $x$ cells of *E. coli* in generation 0 became $2x$ cells of *E. coli* in generation 1. The cells from generation 1 were allowed to divide one further time, still in the medium with normal, unlabelled ammonium ions (Figure 18c). These $4x$ *E. coli* cells constituted generation 2.

DENSITY GRADIENT CENTRIFUGATION

CYTOPLASM

ITQ 2 Assuming DNA does replicate semi-conservatively, what kinds of DNA helix—in terms of heavy and light strands—do you think will be found in generation 2?

Using a technique known as **density gradient centrifugation**, Meselson and Stahl were able to confirm these predictions. The technique of centrifugation in its simplest form was outlined in Unit 22. Density gradient centrifugation is a variation of that simple kind, and can separate DNA molecules that differ only very slightly in their density as a result of the different masses of $^{14}N$ and $^{15}N$. Special centrifuge tubes are used that contain an aqueous solution of a suitable salt such as caesium chloride. The tube is prepared in such a way that there are successive layers of decreasing density of salt solution from the bottom to the top of the tube. A solution of DNA is placed on top of this 'gradient'. If the tube is now centrifuged, a continuous gradient of increasing density of salt will form in the tube, and the DNA molecules will move to the region of the tube where their density equals that of the salt solution and then stop. The position of the DNA in the centrifuge tube can be determined by using ultraviolet radiation of various wavelengths, since DNA solutions absorb strongly in the ultraviolet (u.v.) region of the magnetic spectrum. The method can be made quantitative, so that the amounts of DNA in the different bands can be measured.

Figure 19 summarizes the procedure that Meselson and Stahl used to separate 'light' and 'heavy' DNA. Although this experiment shows only that semi-conservative replication operates in *E. coli*, all the available evidence from a wide range of other experiments indicates that semi-conservative replication of DNA is the universal mode of DNA replication in all organisms that contain double-stranded DNA as their genetic material.

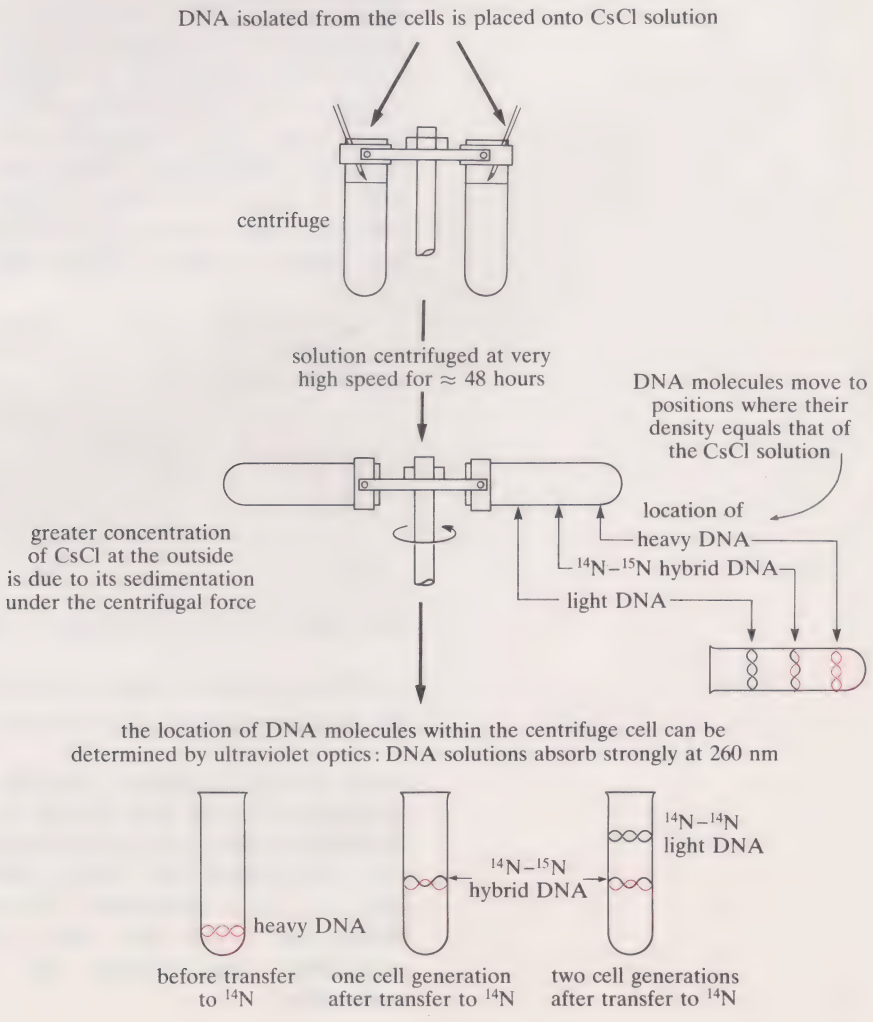



**FIGURE 19** Outline of the procedure Meselson and Stahl used to separate 'heavy', 'light' and 'hybrid' DNA.

At this point, we leave the principles of DNA replication and turn, in Section 6, to Question 4 of the Introduction: 'How does genotype influence phenotype?'

## SUMMARY OF SECTION 5

1 DNA starts to replicate by the two strands separating. Free nucleotides in the nucleus then pair up against the exposed bases of the separated strand (following the base-pairing rules, A–T and C–G). These nucleotides then link together to form two new double helices which are identical to each other *and* to the original double helix. This DNA production process is called semi-conservative replication.

2 That semi-conservative replication takes place is confirmed by experiments using density gradient centrifugation. These show that the original strands of the DNA double helix *are* conserved throughout cycles of DNA replication.

**SAQ 7** Figure 20 shows part of a DNA molecule during replication. A square represents each base, and the number shows its position. Table 2 gives the identities of the bases in positions 3, 6, 7, 9, 11, 20 and 21. Attempt to identify the bases at all other locations in Figure 20 by completing Table 2 as far as you are able. (*Hint:* Draw a sketch of Figure 20, but replace the numbers by A, C, G and T where you can.)

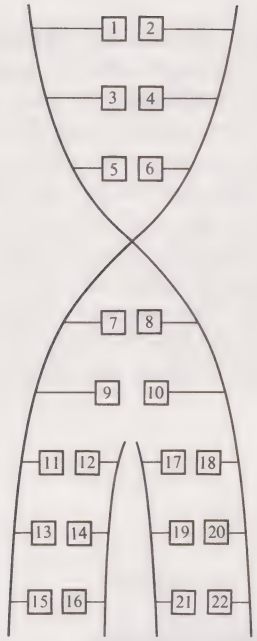



FIGURE 20 Part of a DNA molecule during replication.

TABLE 2 For use with SAQ 7

| Position | Identity | Position | Identity |
|---|---|---|---|
| 1 | | 2 | |
| 3 | adenine | 4 | |
| 5 | | 6 | adenine |
| 7 | guanine | 8 | |
| 9 | guanine | 10 | |
| 11 | cytosine | 12 | |
| 13 | | 14 | |
| 15 | | 16 | |
| 17 | | 18 | |
| 19 | | 20 | adenine |
| 21 | cytosine | 22 | |

**SAQ 8** Reread the description of the Meselson and Stahl experiment in Section 5.2. Suppose the *E. coli* cells of generation 1 had been allowed to divide in medium containing unlabelled (light) ammonium ions, to give generations 2, 3 and 4. In what ratio would the heavier and lighter double helices occur in the cells of generation 4? (Assume that the culture was maintained undisturbed in the same $^{14}NH_4^+$ medium until extraction and centrifugation were carried out at the end of generation 4.)

# 6 PROTEIN SYNTHESIS

How does DNA dicate the structure of protein? Consider the following experimental observations.

When the nucleus of the single-celled marine organism *Acetabularia* (Figure 21) is removed from the cell, the wounded cell heals and continues to make proteins in the cytoplasm for at least two weeks. (**Cytoplasm** is the term used for the cytosol together with everything else inside the cell except the nucleus.)

RIBONUCLEIC ACID, RNA

RIBOSOMES

MESSENGER RNA, mRNA

TRANSFER RNA, tRNA

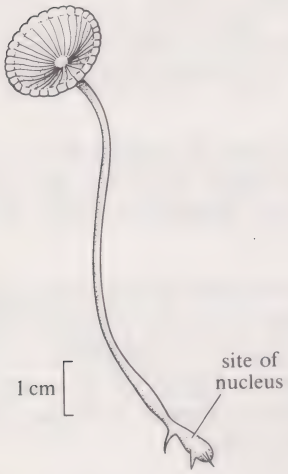



FIGURE 21 *Acetabularia.*

Evidently the nucleus (and its DNA) is not directly involved, in an immediate sense, in continuous protein synthesis. If it were, a cell of *Acetabularia* would stop making protein as soon as its nucleus was removed. It follows that there must be something (linking the DNA in the nucleus and cytoplasmic protein synthesis) that is left behind when the nucleus is excised. Further work has shown that the 'something' is another kind of nucleic acid called **ribonucleic acid (RNA)**. What happens is that nuclear DNA, containing genetic information which determines the proteins that *Acetabularia* should produce, directs the synthesis of a special kind of RNA. This RNA, carrying within itself a copy of the genetic information in the DNA (which remains in the nucleus), travels to structures called the **ribosomes** in the cytoplasm. At the ribosomes, this RNA (called **messenger RNA, mRNA** for short) then directs the synthesis of the particular proteins characteristic of the organism. When the nucleus is removed in this experiment, messenger RNA already in the cytoplasm continues to do its job of directing protein synthesis. After two weeks or so, the messenger RNA starts to break down and, because no more can be made in the absence of the nucleus, protein synthesis stops. This is just what you would expect from the description 'DNA makes RNA makes protein', and in the rest of this Section we shall explore this route further.

It appears that messenger RNA has a similar role in all organisms. In simple terms, we can say that the assembly of each kind of protein is organized at the ribosomes by a specific mRNA molecule, and each type of mRNA molecule is formed according to the instructions of a specific length of DNA. The mRNA molecule moves through pores in the nuclear membrane and travels in the cytoplasm to the ribosomes where proteins are made. Figure 22 (which is familiar to you from the back of your biology Units) shows the location of the nucleus and ribosomes in a schematic cell. Although you have yet to discover the details or to assess the experimental evidence, the overall flow of information is this. The *sequence* of bases in a DNA strand determines the *sequence* of bases in an mRNA molecule, which then somehow determines the *sequence* of amino acids in the protein coded for by the DNA.

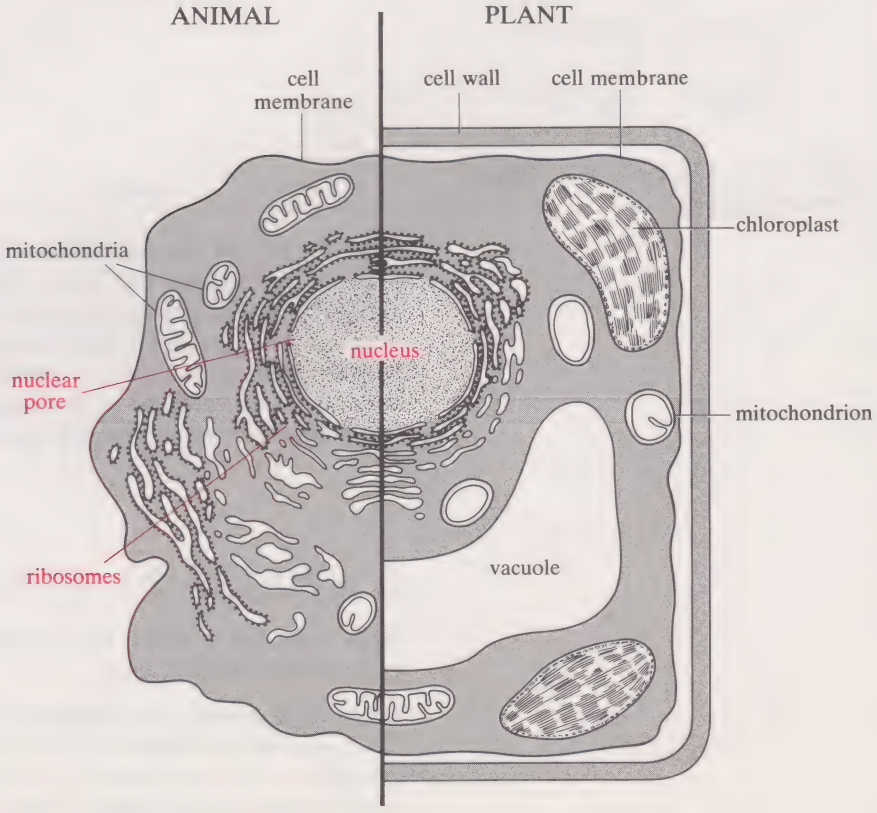



FIGURE 22 Schematic diagram, combining an animal and a plant cell.

There is another kind of RNA involved in protein synthesis called **transfer RNA (tRNA)**. More details about the role of tRNA are given in Section 6.4 but here is a short preview. Suppose part of the genetic information within the nucleus says 'Let the organism produce protein P'. The length of nuclear DNA containing information about P directs the formation of an mRNA molecule in the nucleus, that also contains information about P. That mRNA then travels to a ribosome to which it loosely binds. Next, molecules of the various amino acids involved in the structure of P are collected from the cytoplasm, in which they are dissolved, and delivered at the ribosome by special 'collecting and delivering' molecules—tRNA molecules. There are specific tRNA molecules for each amino acid molecule, and a more or less permanent supply of these continuously resides in the cytosol. Finally, under the direction of the mRNA molecule (still attached to the ribosome), the amino acids are detached from the tRNA carriers and joined together in the specific sequence characteristic of P. The whole process, from DNA to protein P, is summarized in Figure 23.

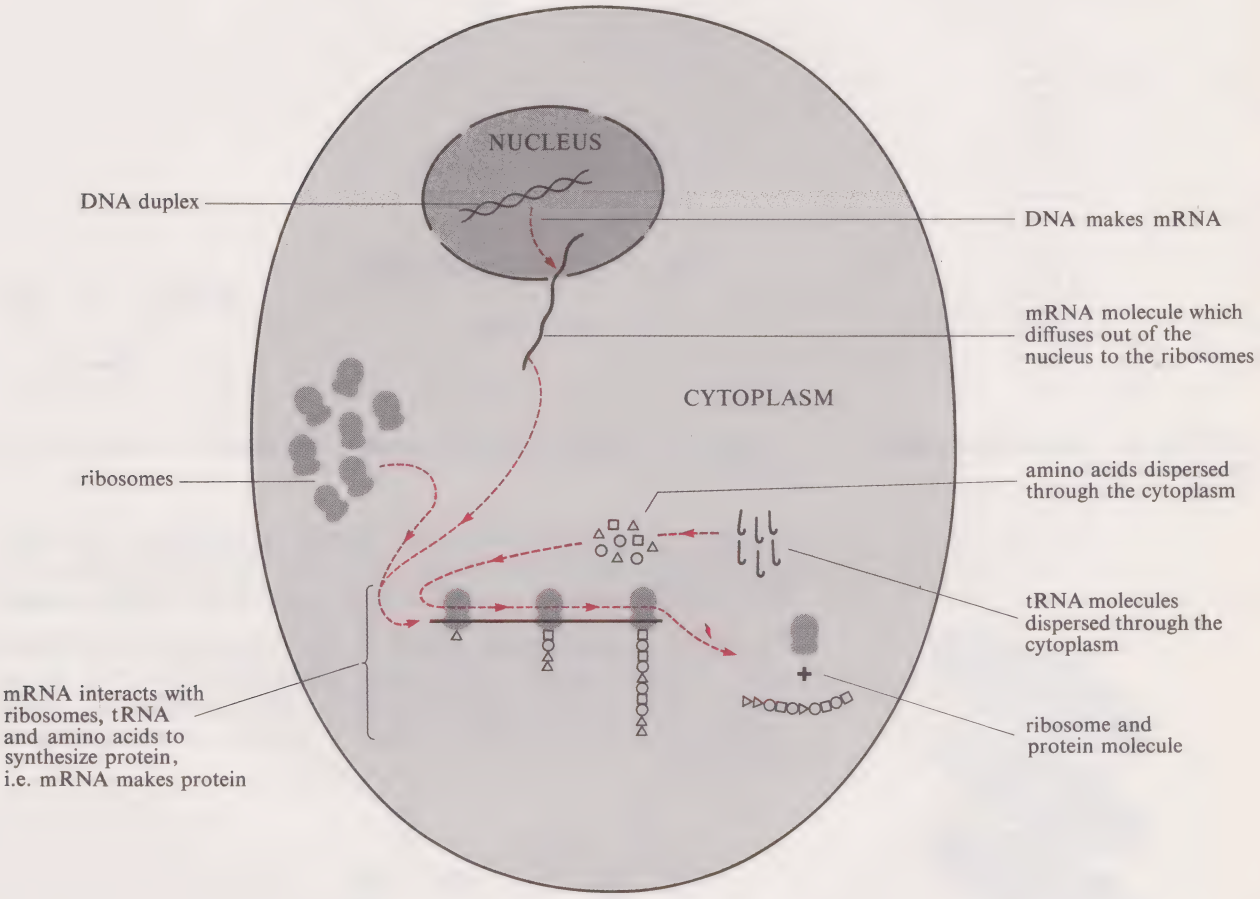



FIGURE 23 A simplified scheme of protein synthesis, showing the two main stages: 'DNA makes mRNA', and 'mRNA makes protein'.

At this stage you may well find this description rather difficult to comprehend! Rest assured it *will* become clear once you have worked through the rest of Section 6.

To understand what is going on in Figure 23, more information is needed about certain aspects of protein synthesis. We need to know:

(a) the structure of the RNA molecules;

(b) how DNA directs the synthesis of RNA;

(c) how the various RNA molecules are involved at the ribosome in the synthesis of particular proteins from amino acids.

We shall examine structure first.

RIBONUCLEOTIDES

POLYRIBONUCLEOTIDE

RIBOSE

URACIL, U

TRANSCRIPTION

## 6.1 THE STRUCTURE OF RNA

Both mRNA and tRNA are *single-stranded* condensation polymers of **ribonucleotides**. We can show their **polyribonucleotide** structure schematically (Figure 24). Each box represents one ribonucleotide.

You will recall from Units 17–18 that a ribonucleotide (sometimes, like deoxyribonucleotide, shortened to nucleotide) is very similar to a deoxyribose in that it consists of three covalently linked parts: a sugar, a phosphate group, and a nitrogen-containing ring called a base (Figure 25). There are four different kinds of base in RNA—adenine (A), guanine (G), cytosine (C) and **uracil (U)**—and it therefore follows that there are four different kinds of ribonucleotide. The bases adenine, guanine and cytosine are identical to those found in DNA.

We can rewrite the structure of RNA in a fuller way (Figure 26). This description probably reminds you strongly of the DNA structure described

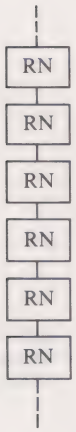



FIGURE 24 The polyribonucleotide structure of RNA.

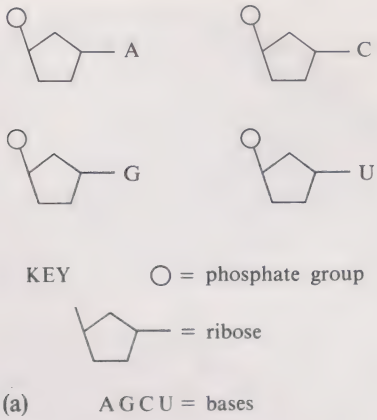



FIGURE 25 (a) The structures of the four ribonucleotides. (b) The full structure of ribose.

in Section 3.2. It certainly should! The only differences are: (a) where DNA has deoxyribose, RNA has **ribose**; (b) where DNA has thymine (T), RNA has uracil (U); (c) RNA is single stranded and DNA is double stranded.

For chemical completeness, Figure 27 shows the structures of thymine and uracil. Thymine has a $-CH_3$ group where uracil has $-H$, but in base pairing terms they are identical. The structures of guanine, cytosine and adenine were given in Figure 9. There is no need to memorize any of these structures.

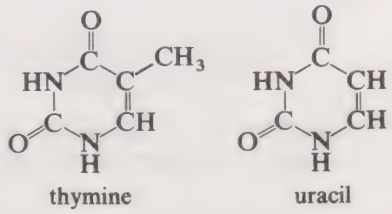



FIGURE 27 The structures of thymine and uracil.

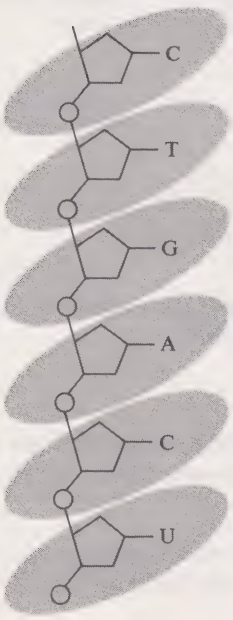



FIGURE 26 The polyribonucleotide structure of RNA shown in more detail than in Figure 24. Each shaded oval is a ribonucleotide. The backbone of the strand consists of alternate phosphate and ribose components.

## 6.2 'DNA MAKES mRNA . . .': THE PROCESS OF TRANSCRIPTION

From Figure 23 it is clear that the genetic information in a DNA molecule inside the nucleus must somehow be passed on to an mRNA molecule, which then moves out of the nucleus to the cytoplasm where protein synthesis occurs. This passing on of the genetic information (i.e. information about which proteins should be made) from DNA to mRNA is called **transcription**. But what is transcription in molecular terms? The answer, confirmed by many experiments, can be appreciated by considering the molecular structures of the DNA double helix and single-stranded mRNA. As noted earlier, it is the *base sequence* in DNA that contains the genetic

information. Thus the DNA base sequence must somehow determine the mRNA base sequence. What happens is this:

1 The strands of the DNA double helix begin to separate from one another, due to the action of specific enzymes.

2 Ribonucleotides, already in solution inside the nucleus, align themselves against *one* of the two exposed DNA strands in accordance with the base-pairing rules shown in Table 3. You can see that these rules are virtually identical to those involved in base pairing during DNA replication (Sections 3 and 5), the only difference being that uracil has replaced thymine in the RNA so that A now pairs with U.

3 As each ribonucleotide arrives at its correct position, it polymerizes with its neighbouring ribonucleotide so that a linear molecule of mRNA is built up.

4 Newly formed mRNA leaves the DNA strand against which it was formed, and the original DNA helix re-forms.

These four stages, which together constitute transcription, are shown in Figure 28. A 'close-up' of stage 3, illustrating base pairing between one DNA strand and the mRNA formed against it, is shown in Figure 29.

TABLE 3 Base-pairing rules in transcription

| DNA base | Ribonucleotide base |
|---|---|
| A | U |
| G | C |
| C | G |
| T | A |

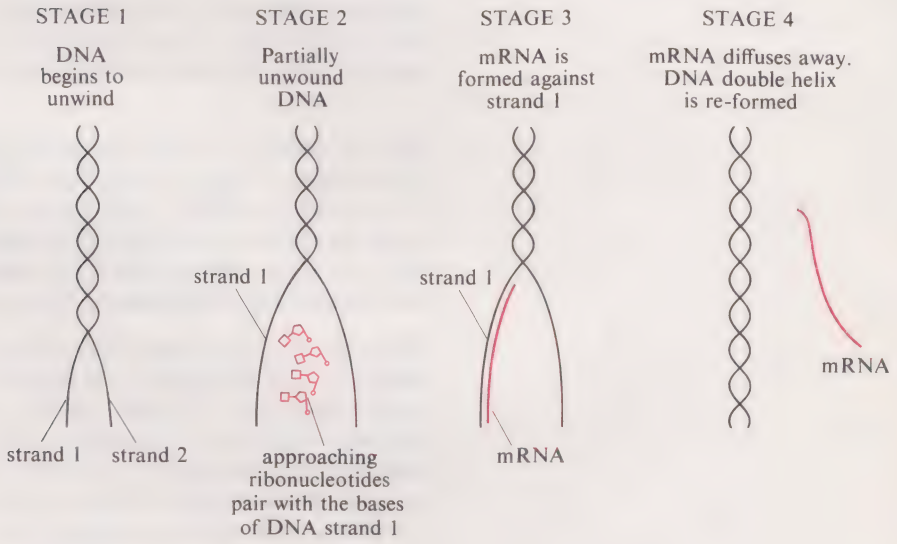



FIGURE 28 The four stages in the process of transcription.

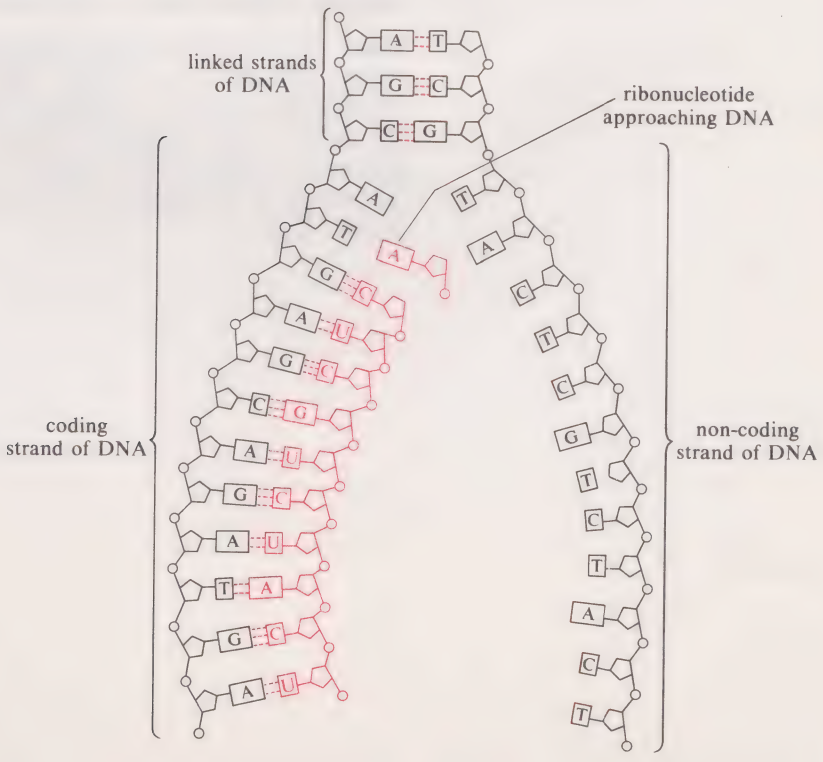



FIGURE 29 Base pairing in transcription; mRNA and hydrogen bonds are shown in red.

GENETIC CODE

TRANSLATION

CODON

You will notice in Figure 28 that only one strand of the DNA is transcribed to make mRNA. The other strand of DNA, although it does not carry information, is nevertheless essential: as you know from Section 5, two strands of DNA are required for semi-conservative replication. Although Figure 28 implies that about half of the DNA is coding for mRNA, the drawing is highly schematic and not to scale. In fact, as you will learn in Section 9, only a fraction of the total length of DNA in a cell is transcribed to make mRNA at any particular time.

A good analogy of the transcription process might be to consider the DNA as a piece of two-stranded and twisted electric flex a metre long. Now imagine that you glue the strands together, but leave a small section, one or two centimetres in length, unglued at some point. If you now hold one end firm and try to untwist the strand, the unglued section of flex will tend to open out. In very simple terms, this is what happens when DNA is transcribed. The double helix maintains the structure of the DNA along most of its length, but a small section unwinds to allow the ribonucleotide bases to align themselves along one strand of DNA and hence enable the transcription of the DNA to mRNA. Now try the following ITQ.

ITQ 3 Near the beginning of Section 6, it was stated that the sequence of bases in a DNA strand determines the sequence of bases in an mRNA molecule (and that this subsequently determines the sequence of amino acids in a protein). Using Figure 29, say in two or three sentences how the transfer of information is brought about in transcription.

Thus, by means of transcription, the information in DNA is preserved in the sequence of bases along the mRNA molecule that it makes. Unlike DNA, however, mRNA is mobile, and after travelling to ribosomes it is able to use the information within it to make a particular protein. Thus mRNA's role is to carry information from nucleus to cytoplasm, rather in the way that a letter carries information from a writer to the recipient.

What can be said about the nature of this information? The detailed answer to this question is the subject of Section 7, which deals with the precise sequence of bases which code for a particular amino acid sequence—known as the **genetic code**. Some points, however, are already clear. Information about the primary structure (the amino acid sequence) of a protein must somehow be contained in the order in which the bases occur in mRNA. Transcription simply entails transferring information from one language of four letters (i.e. the bases A, G, C and T in DNA) into another language of four letters (i.e. the bases A, G, C and U in mRNA).

When the mRNA molecule reaches a ribosome and begins to make a protein, however, we are faced with a more complicated problem. How is information written in a language of *four* letters in mRNA, used to dictate the order (i.e. the sequence) of up to *twenty* amino acids, all of which are different? To answer this question, we need to look a little more closely at the nature of the genetic code and at the details of the 'mRNA makes protein' step.

## 6.3 '. . . mRNA MAKES PROTEIN': THE PROCESS OF TRANSLATION

Refer back to Figure 23. As you can see, the second and final stage of protein synthesis involves the interaction of an mRNA molecule, loosely attached to a ribosome, with many amino acid molecules. (Remember that although there are only 20 different kinds of amino acid, one molecule of a particular protein may contain many amino acids of each type.) Each amino acid molecule is brought to the ribosome by its own tRNA molecule (this is discussed in Section 6.4). The result of this interaction is the production of a protein molecule in which the constituent amino acids are linked together in *exactly the right sequence*; that is, the sequence dictated by the mRNA and uniquely characteristic of the particular protein. The whole set of events is termed **translation**. The 'message' in the mRNA is translated to form a protein molecule. How does this occur? Any attempt at an answer must begin with a closer look at how mRNA contains information about protein structure. It is here that you will begin to realize why the information in mRNA (and also in DNA) is said to be in *code*. Just four different bases in mRNA have the task of coding for the type and position of the 20 different amino acids in a protein. This is rather like (but do not push the analogy too far) the sequence of two symbols in the Morse code (dot and dash) that are used to code for the 26 letters of our ordinary alphabet. In fact—as is explained more fully in Section 7—a group of *three* adjacent bases in mRNA (and hence in DNA) represents *one* amino acid. Each of these sets of three bases is called a **codon**, and it is the *sequence* of the codons that dictates the sequence of amino acids in the protein.

□ As you will see later in the Unit, the codon UUU in the mRNA codes for the amino acid phenylalanine, the codon GGA codes for glycine, and the codon UCG codes for serine. What would be the sequence of amino acids within a protein if the mRNA had a base sequence:

(a) UCG GGA UUU;

(b) GGA GGA GGA?

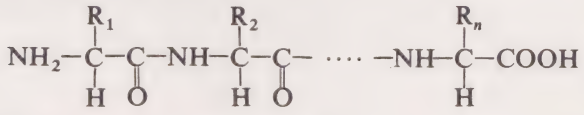
■ These sequences of bases would give the following amino acid sequences:

(a) serine–glycine–phenylalanine;

(b) glycine–glycine–glycine.

Because the stage 'mRNA makes protein' involves deciphering the code (i.e. going from a language using four letters to a language using 20 letters), the term *translation* is very appropriate.

Now let us leave coding for the time being and consider what happens at ribosomes. Remember from Unit 22 that, when amino acids condense to form proteins, the $-\overset{\overset{O}{\|}}{C}-OH$ of one amino acid links with the $-NH_2$ of an adjacent amino acid to form a peptide bond, with the elimination of water. Specific enzymes are involved in catalysing this reaction. Thus a protein can be represented as:



For the purposes of this Unit, all you need note is that one end of a protein molecule has an amino acid with a free $-NH_2$ group and the other end has an amino acid with a free $-COOH$ group. As you learnt in Unit 22, these

N-TERMINAL AMINO ACID

C-TERMINAL AMINO ACID

STOP CODON

POLYSOME

are called the **N-terminal amino acid** and the **C-terminal amino acid**, respectively. When a protein is made at a ribosome, it is built up step by step, one amino acid at a time, beginning at the N-terminal end and finishing at the C-terminal end. Conventionally, the process is regarded as taking place from left to right.

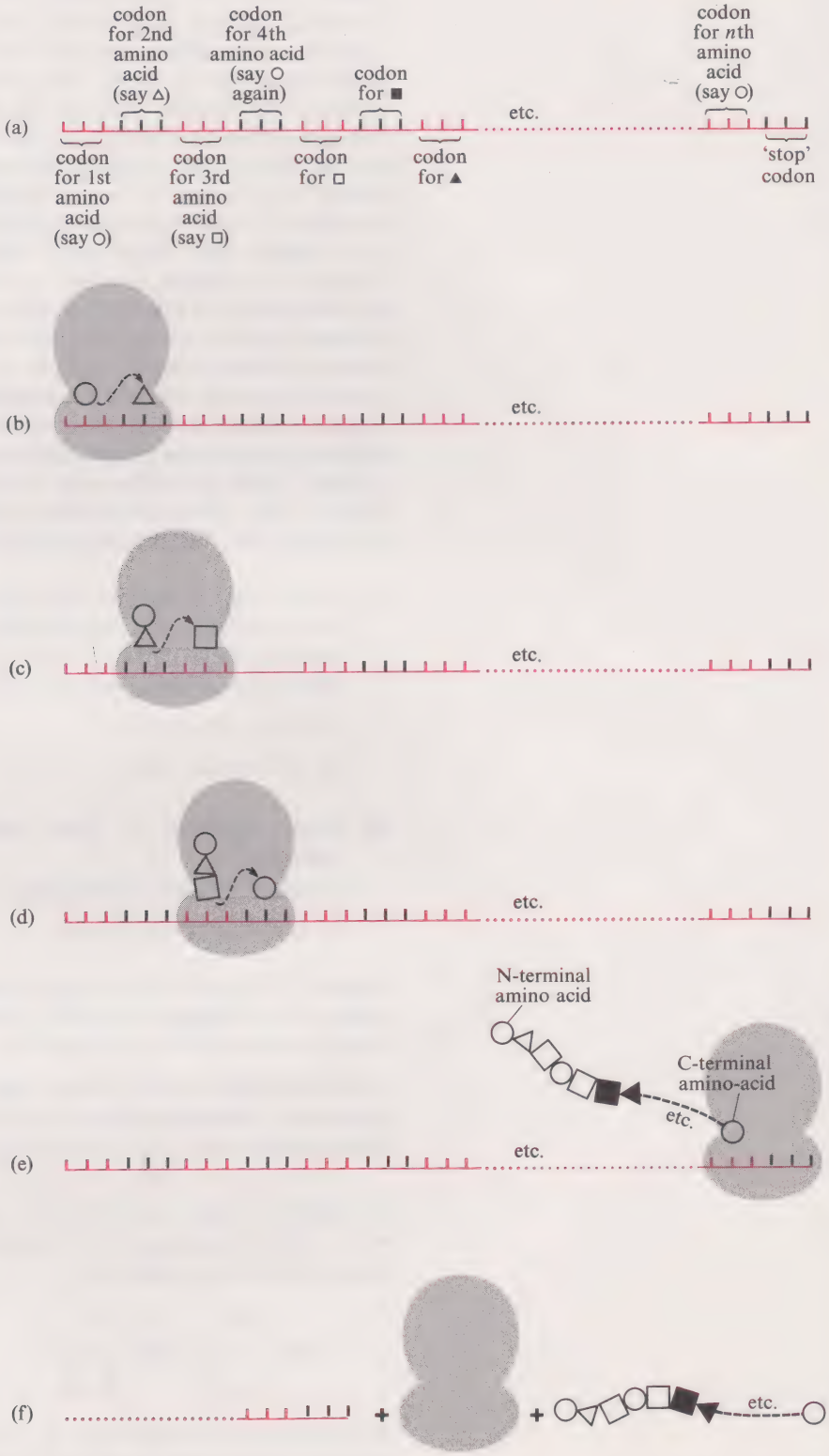



FIGURE 30 A summary of translation. For details, see text.

The whole process is summarized in Figure 30. (For the moment we are ignoring the role of tRNA.) Working through this Figure, you can see how a polypeptide chain is built up on the mRNA template:

(a) An mRNA molecule, built up from many codons, moves from the nucleus to the cytoplasm. (Alternate codons are shown in red, for clarity.)

(b) A ribosome attaches itself to the far left-hand end of the mRNA molecule. The two codons (i.e. triplets of bases) nearest that end cause two particular amino acids, namely the two coded for by those codons, to position themselves on the ribosome above the codons. The first amino acid—the N–terminal amino acid—then moves and links itself to the second amino acid to give a dipeptide. This condensation reaction involves enzymes and ATP, the latter being converted to ADP and $P_i$.

(c) Immediately afterwards, the ribosome moves to the right along the mRNA molecule and positions itself over the second and third codons, thus allowing the third amino acid (whose identity is determined by the third codon) to align itself over that codon. The dipeptide now moves and condenses with the third amino acid to give a tripeptide.

(d) Immediately following this condensation, the ribosome once again moves to the right, this time positioning itself over the third and fourth codons. Thus the fourth amino acid (whose identity is determined by the fourth codon) now aligns itself over that codon. The tripeptide now moves and condenses with the fourth amino acid to give a tetrapeptide, thus building up the next step in the polypeptide chain.

(e) The ribosome continues to pass along, codon by codon, from left to right, bearing a longer and longer peptide chain. At each step a further amino acid is added; each time, this involves catalysis by appropriate enzymes and the conversion of ATP to ADP. When the ribosome reaches the end of the mRNA molecule, it meets what is called a **stop codon**. This terminates polypeptide chain synthesis—the last amino acid added before the stop codon being the C-terminal amino acid.

(f) The polypeptide and the ribosome are set free, leaving the mRNA molecule unencumbered. Both the ribosome and the mRNA molecule are reusable: each mRNA molecule normally directs the synthesis of many molecules of a particular polypeptide.

You may well find Figure 30 easier to understand and remember than the text! In some ways, the situation can be likened to a self-service cafeteria: the plate (the ribosome) goes from left to right past a series of serving points (each mRNA codon), receiving the appropriate item (each amino acid). At the end of the sequence the plate is full (the polypeptide is complete and released).

The term **polysome** is used to describe a working group of ribosomes: each polysome is a group of ribosomes that are all attached to a single thread of mRNA. What happens in a polysome is that a number of different ribosomes work their way along the same mRNA thread, each building up its protein chain as it goes. Those nearest the right-hand end have nearly complete protein chains. Each ribosome in the polysome builds up the same kind of protein molecule (i.e. the one coded for by the mRNA). On reaching the stop codon at the far right-hand end, each ribosome and its completed protein pass into the cytoplasm (see Figure 31).

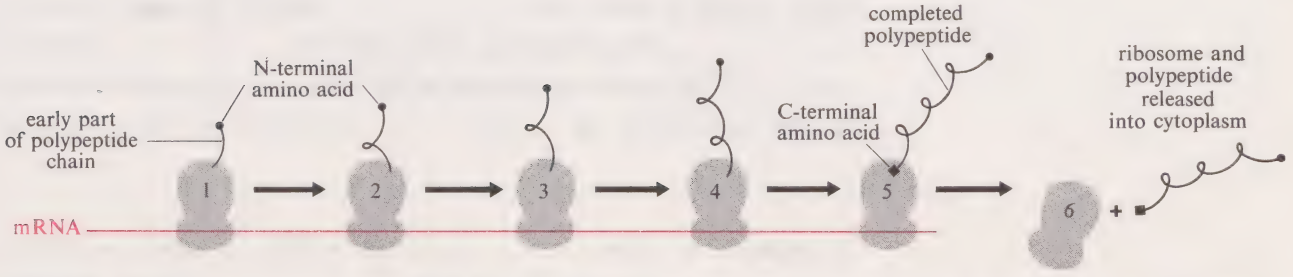



FIGURE 31 A polysome consisting of an mRNA molecule and five different ribosomes. The arrows show the direction of movement of each ribosome.

AMINO ACID–tRNA COMPLEX

tRNA ANTICODON

ANTICODON

To put these diagrams into perspective, have a look at Figure 32. This is an electron micrograph of some polysomes involved in protein synthesis in *E. coli*. A number of ribosomes are arranged along a length of mRNA to form each polysome. The faint line in the photograph is the bacterial DNA. Unlike eukaryotic cells, in *E. coli* the mRNA is translated immediately after transcription. This means that it is still attached to the DNA when translated, and is the reason why the polysomes in this Figure appear to be attached to the DNA molecule.

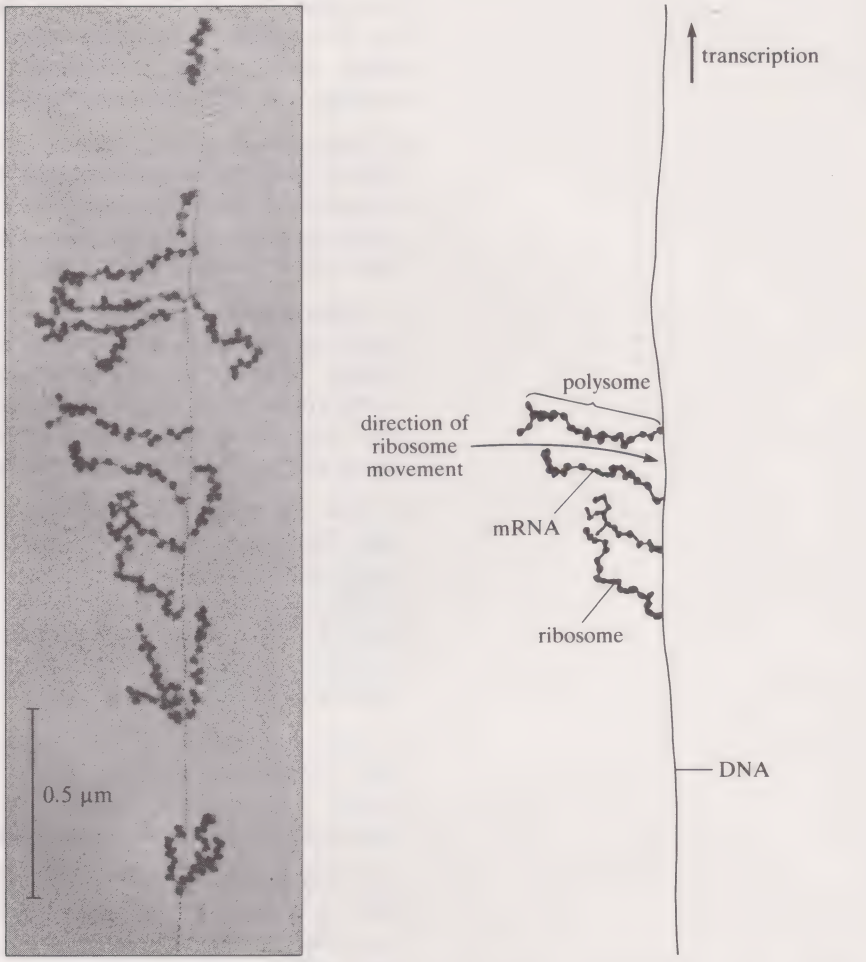



FIGURE 32 Electron micrograph of DNA, mRNA and polysomes in *E. coli*, with a key to the structures on the right.

You should now work through ITQ 4 before continuing with this Section. Don't be tempted to skip it!

**ITQ 4** Fill in the missing words in the following paragraphs, which outline the process of translation. You will need to read several sentences ahead of each blank before you can fill it in with confidence.

DNA carries a code for ..... After a process known as ....., these molecules move into the ..... of the cell via pores in the nuclear membrane. In the cytoplasm, they act as templates for the production of ..... molecules by the process of .....

A sequence of three ..... in mRNA is called an mRNA ..... The sequence of these ..... along a particular

mRNA molecule determines the specific sequence of ..... ..... in the ..... coded for by the..... of the gene and its complementary .....

Next, in a progressive manner, these ..... ..... position themselves over the appropriate mRNA codons. At each step, through the catalytic action of specific ..... and with conversion of ..... to ..... and $P_i$, an amino acid is added to the growing ..... chain.

Ultimately, when the ..... codon is reached, the completed ..... molecule is released into the cytoplasm. Several ..... are able to move along one mRNA molecule, each bearing a progressively longer ..... chain. The overall structure—consisting of an mRNA molecule, several ....., and ..... chains of different lengths—is called a .....

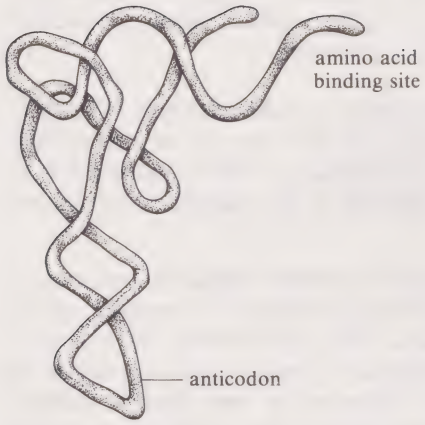



FIGURE 33 The structure of a tRNA molecule.

## 6.4 THE ROLE OF tRNA IN TRANSLATION

Now that we have built up a picture of the general process of translation, we can go back and see how the tRNA molecules fit in. The assumption in Section 6.3 was that amino acids are somehow matched up to their corresponding mRNA codons, and that tRNA is involved in some way. It has been demonstrated experimentally that each tRNA molecule links a particular amino acid to its appropriate codon. Figure 33 shows the three-dimensional structure of a typical tRNA molecule.

There are 20 different amino acids, so there are (at least) 20 different tRNA molecules. Each tRNA molecule is made in the nucleus, from a specific sequence of DNA bases, in a similar way to mRNA. A particular kind of tRNA will form a complex with only one kind of amino acid; for example, $tRNA_{Ala}$ will form a complex only with alanine, and $tRNA_{Gly}$ will form a complex only with glycine. How does a tRNA bind its own amino acid, and how does the **amino acid–tRNA complex** find its way to the appropriate mRNA codon?

1 The first step depends on specific activating enzymes. For example, the activating enzyme for glycine, $enzyme_{Gly}$, catalyses the reaction:

$$\text{glycine} + tRNA_{Gly} \xrightarrow{enzyme_{Gly}} \text{glycine–}tRNA_{Gly}$$

2 The second step depends on the fact that each tRNA molecule possesses a specific site that recognizes and binds to the appropriate mRNA codon. This site is called the **tRNA anticodon** (or just the **anticodon**) and consists of three unpaired bases which are complementary to an mRNA codon. In this way, $tRNA_{Gly}$, bearing its complexed glycine molecule, aligns itself with a glycine codon on the mRNA molecule. Figure 34 shows, in very simplified form, a glycine–$tRNA_{Gly}$ complex paired with a glycine codon on the mRNA. We shall use this simplified 'walking stick' representation of tRNA in future diagrams.

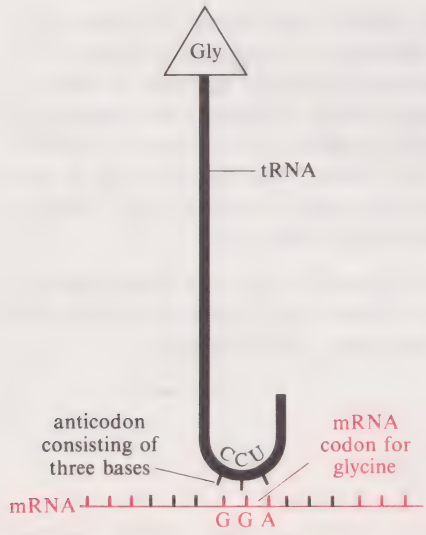



FIGURE 34 Base pairing between a tRNA anticodon (part of an amino acid–tRNA complex, glycine–$tRNA_{Gly}$) and complementary bases on the mRNA.

☐ What sort of forces do you think will hold the tRNA in place on the mRNA?

■ The forces associated with hydrogen bonds—the same kind of forces as hold the two strands of DNA together.

As you know, hydrogen bonds are weaker than covalent bonds (Units 17–18). They are just strong enough to hold the amino acid–tRNA complex in position for the amount of time (estimated to be about 0.01 second) necessary for the condensation reaction to occur, but weak enough to

RIBOSOMAL RNA, rRNA

permit the 'spent' tRNA molecule to escape from the mRNA surface once the amino acids have been joined together. The ribosome plays an important part here—in effect it provides a protected environment where these weak interactions can proceed undisturbed for the short time necessary for the strong peptide bonds to form. You could think of the ribosome as a ring doughnut through which the mRNA is threaded, with the hole in the doughnut providing this protected environment.

## 6.5 TRANSLATION REVISITED

It will be apparent to you that the description of translation in Section 6.3, and especially in Figure 30, is much simplified! We shall now look in more detail at the roles of first the ribosomes, and then the tRNA.

Ribosomes are small particles made of protein plus yet another kind of RNA—**ribosomal RNA (rRNA)**. Although rRNA does not carry *information* in the very specific way that mRNA does, its role in protein synthesis is nevertheless crucial. It is synthesized in a similar way to mRNA.

□ Where do you think rRNA is made?

■ In the nucleus, where the other RNA molecules are also made. There are specific sequences of DNA which carry the code that is transcribed to form rRNA molecules. (A structure in the nucleus, called the nucleolus, is thought to be associated with its production.)

A given ribosome can bind, at different times, to different types of mRNA molecule. Thus, ribosomes have nothing to do with the *type* of protein that is formed. rRNA plays a major role in the binding of tRNA and mRNA at the ribosome, but exactly how it is involved is the subject of continuing research.

We now need to tell you more about tRNA. tRNA molecules are, as you have seen, essential in the process of protein synthesis. In carrying out its role of adaptor molecule between amino acids and mRNA codons, each tRNA molecule has *two* kinds of recognition power: it can recognize its own specific amino acid, and it can recognize its own specific mRNA codon. All this implies a unique structure for each kind of tRNA. It will be no surprise to you, therefore, that each kind of tRNA is synthesized when necessary at its own specific gene within the nuclear DNA, to top up the supply of tRNA in the cytosol.

Figure 35 gives the essential points of tRNA involvement in translation, showing the stages (b), (c) and the first moments of stage (d) of Figure 30 in more detail. As an illustration, real identities have been ascribed to the first, second and third mRNA codons, though totally different arrangements of the amino acids would have been equally possible. Note that each tRNA molecule can be used and re-used many times. You can also see in this Figure that the ribosome has two amino acid–tRNA binding sites, P and A, which lie on the left- and right-hand side of each ribosome.

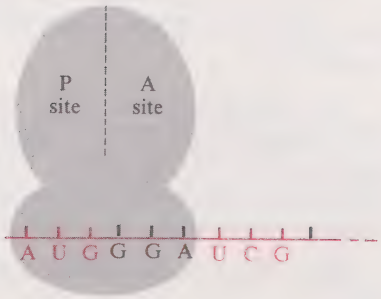
The most important point to note from Figure 35 is the role of each specific tRNA molecule in aligning a particular amino acid (alone or as part of a growing polypeptide chain) against a particular mRNA codon.

## SUMMARY OF SECTION 6

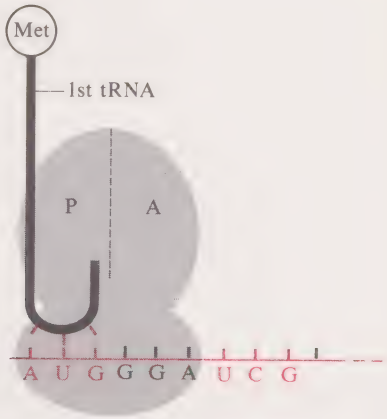
Protein synthesis in all cells which have DNA as their genetic material follows a basic pattern:

1 DNA is transcribed to form three different kinds of RNA molecule: mRNA, tRNA and rRNA. The base pairing rules are the same as for DNA replication, except that—since T is replaced by U in RNA—A pairs with U.

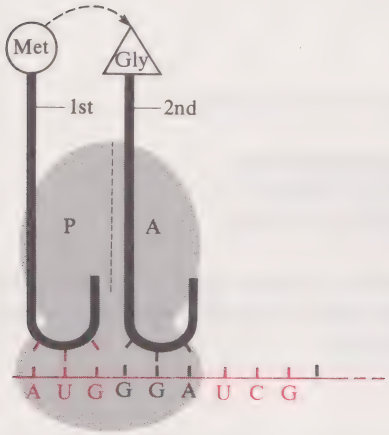

1 The left-hand end of part of an mRNA molecule, unattached to a ribosome



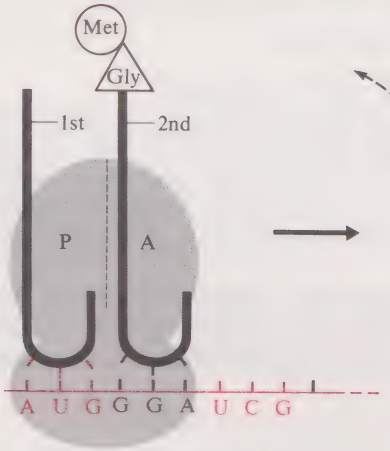

2 A ribosome attaches to the mRNA



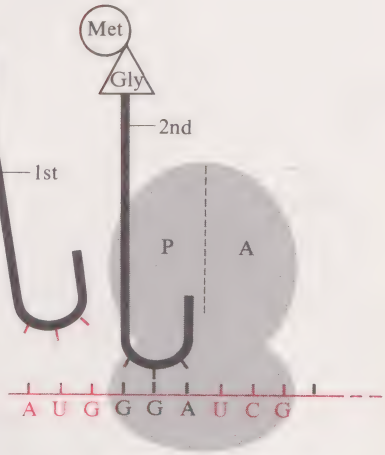

3 The N-terminal amino acid methionine attached to its tRNA binds to the P site of the ribosome. The tRNA anticodon pairs with AUG (mRNA codon for methionine)



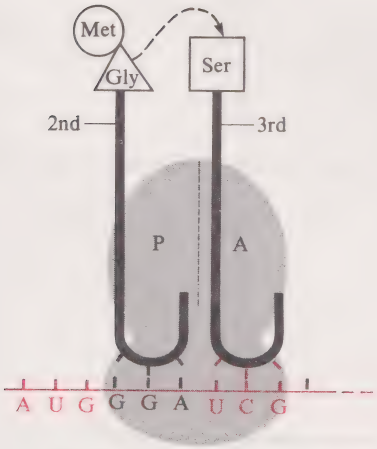

4 glycine–tRNA binds to the A site. Then the N-terminal amino acid migrates to form a dipeptide



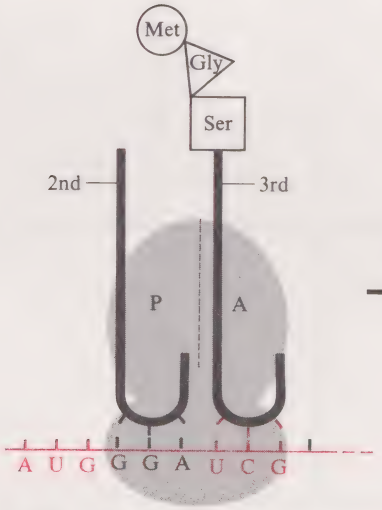

5 Ribosome movement now occurs



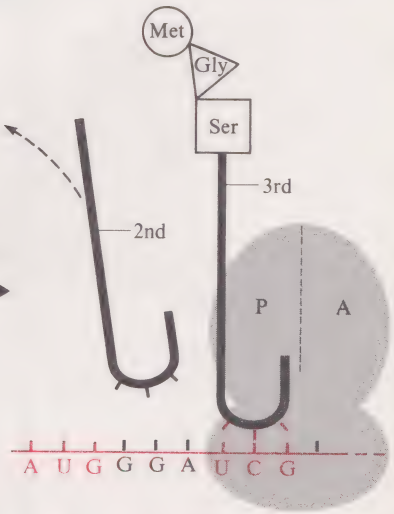

6 The unloaded tRNA molecule leaves the mRNA. The empty A site is ready to receive the next tRNA–amino acid complex




7 The dipeptide migrates to form a tripeptide




8 Ribosome movement now occurs




9 The A site is once again empty; the unloaded tRNA molecule leaves the mRNA. This is a repeat of stage 6

FIGURE 35 A more detailed representation of the early translation stage shown in Figure 30. Note that AUG is the mRNA codon for methionine, GGA for glycine, and UCG for serine.

## TRIPLET CODE

A molecule of mRNA carries the code for a particular protein. tRNA and rRNA have different roles in the process of translation.

2 The rRNA associates with proteins to form ribosomes.

3 mRNA binds to ribosomes in the cytoplasm.

4 Specific amino acid–tRNA complexes are formed in the cytoplasm.

5 The ribosome provides a protected environment for the mRNA, where these amino acid–tRNA complexes bind to specific sequences of bases along the mRNA.

6 The ribosome moves along the mRNA, thus allowing a sequence of tRNA molecules (with their amino acids) to pair up at the correct position against codons on the mRNA. Each successive amino acid condenses with the growing peptide chain, eventually building up the complete polypeptide. At each condensation step, enzymes and ATP are involved.

7 Stop codons on the mRNA indicate when the polypeptide chain is complete.

**SAQ 9** Which of the descriptions (i)–(vi) apply to:

(a) both DNA and RNA;
(b) RNA but not to DNA;
(c) DNA but not to RNA?

*Descriptions*

(i) linked nucleotides
(ii) typically found in the nucleus but not in the cytoplasm
(iii) typically found as single strands
(iv) contain(s) as many purine bases as pyrimidine bases
(v) contain(s) nucleotides in which the sugar is ribose
(vi) contain(s) nucleotides, some of which contain uracil

**SAQ 10** Complete Table 4 to show where the three types of nucleic acid involved in protein synthesis in eukaryotic cells are formed. Also show their function(s), and where the function(s) take place.

TABLE 4 For use with SAQ 10

| Type of nucleic acid | Where formed | Function(s) | Where function(s) takes place |
|---|---|---|---|
| DNA | | | inside the nucleus |
| mRNA | | | |
| tRNA | | forms complexes with specific amino acids, then binds to specific mRNA codons | |

**SAQ 11** Which of the following are TRUE statements and which are FALSE? Explain your answers briefly.

(a) The translation of genetic information in cells involves mRNA, tRNA and ribosomes.

(b) After transcription, ribosomes become detached from the sequence of DNA involved in protein synthesis.

(c) Each amino acid attaches directly to its own specific mRNA codon.

(d) mRNA is synthesized inside the nucleus of eukaryotic cells and passes into the surrounding cytoplasm where it takes part in protein synthesis.

(e) Each mRNA molecule can be used only once for synthesizing protein.

(f) Polypeptides are synthesized step by step, starting from the C-terminal amino acids.

(g) Transcription of genetic information takes place in the nucleus of eukaryotic cells.

(h) Genetic information is translated in the cell nucleus.

# 7 THE GENETIC CODE

If tRNA is like mRNA in containing only the four bases A, G, C and U, it is possible to deduce tRNA anticodons from mRNA codons. Figure 36 summarizes these relationships for some of the amino acids whose DNA codons were introduced in Section 6. (Note that the non-coding strand of DNA has a sequence of bases similar to that of mRNA—except that T in the DNA is replaced by U in the mRNA.)

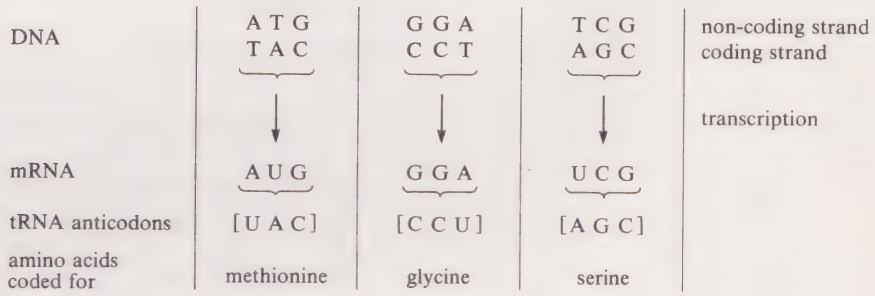



FIGURE 36 The relationship between DNA codons (upper brackets), mRNA codons (lower brackets) and the anti-codons found on each tRNA molecule (square brackets).

Figure 36 represents only a small part of what is now known about the genetic code. These details were discovered between the mid-1950s and the mid-1960s. During this period, molecular biologists demonstrated experimentally that the genetic code is truly a triplet code, did experiments that led to the complete deciphering of the code, and produced evidence supporting the view that the genetic code is universal (that is, it is the same for all organisms). We develop these points in the rest of this Section.

## 7.1 THE TRIPLET NATURE OF THE GENETIC CODE

It is now clear that the genetic code is a **triplet code**. The first evidence for this was purely theoretical. Knowing that mRNA contains just four kinds of base and that 20 types of amino acid must be encoded, what can be deduced about the number of bases in a codon? Fairly obviously, the answer cannot be that each codon consists of one base. If it were, only four kinds of amino acid could be encoded. What if codons contained two bases? There are 16 ($4 \times 4$, i.e. $4^2$) different pairs of bases that can be arranged from four different bases: you can check this by examining Table 5. (*Note:* Do not attempt to learn Tables 5 and 6.)

TABLE 5 A 4 × 4 grid, showing the 16 different pairs of letters that can be constructed from four different letters

| | *First letter* U | C | A | G | |
|---|---|---|---|---|---|
| | UU | CU | AU | GU | U |
| | UC | CC | AC | GC | C |
| *Second letter* | UA | CA | AA | GA | A |
| | UG | CG | AG | GG | G |

CELL-FREE SYSTEM

DEGENERATE CODE

WOBBLE HYPOTHESIS

However, this number of possibilities is insufficient to code for 20 amino acids. If we suppose that each codon contains three bases, the result is rather different! There are 64 ($4 \times 4 \times 4$, i.e. $4^3$) different triplets that can be arranged from four different bases. You can check this by glancing forward to Table 6. This total is more than sufficient to encode 20 amino acids: indeed there are plenty left over. Clearly, codons of four, five or more bases are also theoretical possibilities (there are 256, i.e. $4^4$, ways of arranging four different bases in groups of four). However, the smallest theoretically possible codon size is three bases, and a number of experiments (not described here) have shown this to be the actual number; this leads to 64 different triplets, as explained above.

## 7.2 DECIPHERING THE CODE

Table 6 shows the meaning of each of the 64 mRNA codons. (Table 6 is essentially the same as Table 5, repeated four times, but with U placed in front in the first four rows, C in the second four, A in the third and G in the fourth, to allow for the extra variation that the third base provides.) Most descriptions of the genetic code are given in terms of mRNA codons rather than DNA codons, because most of the experiments that have led to the deciphering of the code have involved mRNA and not the inaccessible DNA. These experiments were often technically ingenious but conceptually simple. Work done in 1961 by Heinrich Matthaei and Marshall Nirenberg in the USA provides a glimpse of one approach. They prepared a synthetic mRNA termed poly(U), containing uracil (U) as the only base. When they added poly(U), together with all 20 amino acids, to a **cell-free system** capable of synthesizing protein (i.e. ribosomes, tRNA molecules, enzymes, ATP, etc.), only one polypeptide, polyphenylalanine, was formed. Thus UUU UUU UUU . . . gave rise to Phe–Phe–Phe . . . , so it was clear that the mRNA codon UUU codes for phenylalanine. Similarly, poly(C)—containing cytosine as the only base—yielded polyproline: hence CCC must code for proline.

TABLE 6 mRNA codons

| *First letter* | *Second letter* U | *Second letter* C | *Second letter* A | *Second letter* G | *Third letter* |
|---|---|---|---|---|---|
| U | UUU, UUC **Phe**; UUA, UUG **Leu** | UCU, UCC, UCA, UCG **Ser** | UAU, UAC **Tyr**; UAA stop; UAG stop | UGU, UGC **Cys**; UGA stop; UGG **Trp** | U C A G |
| C | CUU, CUC, CUA, CUG **Leu** | CCU, CCC, CCA, CCG **Pro** | CAU, CAC **His**; CAA, CAG **Gln** | CGU, CGC, CGA, CGG **Arg** | U C A G |
| A | AUU, AUC, AUA **Ileu**; AUG **Met** | ACU, ACC, ACA, ACG **Thr** | AAU, AAC **Asn**; AAA, AAG **Lys** | AGU, AGC **Ser**; AGA, AGG **Arg** | U C A G |
| G | GUU, GUC, GUA, GUG **Val** | GCU, GCC, GCA, GCG **Ala** | GAU, GAC **Asp**; GAA, GAG **Glu** | GGU, GGC, GGA, GGG **Gly** | U C A G |

The abbreviated names of amino acids are as follows: **Ala** = alanine, **Arg** = arginine, **Asn** = asparagine, **Asp** = aspartic acid, **Cys** = cysteine, **Gln** = glutamine, **Glu** = glutamic acid, **Gly** = glycine, **His** = histidine, **Ileu** = isoleucine, **Leu** = leucine, **Lys** = lysine, **Met** = methionine, **Phe** = phenylalanine, **Pro** = proline, **Ser** = serine, **Thr** = threonine, **Trp** = tryptophan, **Tyr** = tyrosine, **Val** = valine. You do not need to remember these abbreviations.

Scientists were also able to manufacture synthetic polyribonucleotides of a known sequence. It proved possible to make a polymer with an *alternating* sequence of mRNA bases: UGU GUG UGU GUG UGU G . . . . When this polymer was put into a cell-free system (as in Matthaei and Nirenberg's work), it was discovered that the polypeptide formed had a repeating, alternate sequence of the two amino acids cysteine and valine. Thus the codons for these two amino acids were UGU and GUG. Previous work had already confirmed the codon for valine as GUG, thus the codon for cysteine was UGU.

□ Suppose you were able to make a synthetic mRNA with the base sequence AGA GAG AGA GAG AGA G . . . . Referring to Table 6, what sequence of amino acids would you expect to be made from this mRNA template?

■ There are only two possible codons here—AGA and GAG. As you see from Table 6, these code for the amino acids arginine and glutamic acid. Thus you would expect a polypeptide to form which had an alternating sequence of arginine and glutamic acid.

By the end of 1966, experiments such as these had resolved the genetic code as shown in Table 6. There are a number of points about this Table that should be noted:

1 In most cases, one amino acid is coded for by several different codons (perhaps not surprisingly, given that there are 64 triplets and only 20 amino acids). Because of this, the genetic code is described as a **degenerate code**. You can see from the Table that tryptophan (Trp) and methionine (Met) are the only amino acids to have just one codon each. In contrast, leucine (Leu), serine (Ser) and arginine (Arg) each have six different codons. Other amino acids have two, three or four codons each. It is important to note, on the other hand, that each *codon* codes for only *one* amino acid.

2 There are a total of 61 different codons coding for amino acids—if you wish, you can check this by counting them up in Table 6, where there are 64 codons but three (UAA, UAG and UGA) do not code for amino acids. Does this mean that there are 61 different tRNA molecules, each with its own anticodon complementary to a codon? Another way of putting this question is to ask whether you would expect, for instance, four different tRNAs for proline or valine, since each of these amino acids is coded by four different codons—see Table 6. Although there is debate about this issue, there is some evidence that the base pairing between an mRNA codon and a tRNA anticodon need not be as precise as the normal base-pairing rules would seem to demand. Under certain conditions, it may be possible to get unusual base pairings. For instance, G to U as well as A to U, or A to C as well as G to C. This idea, known as the **wobble hypothesis** and first put forward by Crick in 1966, goes a long way towards explaining the patterns in the genetic code, where a number of amino acids can be uniquely defined by just the *first two* bases of the codon. The precise nature of the third base (e.g. proline) is not so crucial. You can see from Table 6 that where two amino acids share the first two bases in the codon (e.g. histidine and glutamine), the third base for each codon is either a 'large' purine or a smaller pyrimidine. The *precise* structure of this third base does not appear to be crucial—hence the codon can 'wobble'. In principle, therefore, we should not be surprised if we find less than 61 different tRNA molecules. It appears that in some organisms as few as 40 different tRNA molecules are present.

3 Three of the 64 codons do not code for any amino acid: these are UAA, UAG and UGA. They appear to function as stop codons within mRNA. As you know from Figure 30, a ribosome travels from left to right (by convention) along an mRNA molecule, carrying with it the growing polypeptide chain. When the ribosome reaches a stop codon, no more amino acid is incorporated; instead, the polypeptide is released from the ribosome as a completed protein.

NON-OVERLAPPING CODE

INITIATOR CODON

FUNCTIONAL PROTEIN

POST-TRANSLATIONAL MODIFICATION

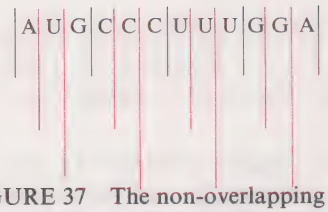



FIGURE 37 The non-overlapping nature of the genetic code. The ribosome 'sees' the triplets AUG, CCC, UUU and GGA (picked out by pairs of black lines), and does not see the overlapping triplets UGC etc. and GCC etc. (picked out by pairs of red lines of the same length).

4 One point that is implicit in Table 6 and in the whole of Section 6 is that **the genetic code does not overlap**: triplets are read from the far left-hand end of mRNA, in groups of three. Thus in Figure 37, the short sequence of mRNA codes for the amino acids represented by the short black lines only. Possible overlapping triplets (shown by the longer red lines) are not 'seen' as codons by the travelling ribosome. We should add, for the sake of accuracy, that there is evidence that a few viruses *do* use an overlapping code. However, in the vast majority of living organisms it appears that the genetic code is non-overlapping.

5 Finally, it appears that not only is AUG the codon for methionine, it is also the codon that initiates *all* polypeptides—and is therefore called the **initiator codon**. It follows from this, of course, that every freshly completed polypeptide chain released from polysomes has methionine in the N-terminal position. In contrast, most **functional proteins** (that is, the *biologically active* protein molecules) do not have methionine as the N-terminal amino acid. It therefore appears that the methionine is removed from the left-hand end of the polypeptide chain after the chain leaves the ribosome. Because this alteration to the chain occurs just after the translational process—and, indeed, is necessary before the protein molecule can exhibit biological activity—the phenomenon is termed **post-translational modification**. You will meet modifications of this type is Section 9, when we consider the biosynthesis of the hormone insulin.

## 7.3 UNIVERSAL NATURE OF THE GENETIC CODE

Before we finish this Section, it is important to emphasize that the genetic code is virtually *universal*. In other words, practically all organisms use the same basic mechanisms for getting from DNA to the functional protein. Consider the following experimental observations:

(a) Synthetic RNA polymers code in the same way (i.e. produce the same polypeptides), whether they are used in mammalian or bacterial cell-free systems. This means that as well as bacteria and mammals having the same mechanisms for protein synthesis—ribosomes, tRNA, enzymes and so on—the specific codes for amino acids must be identical.

(b) mRNA extracted from mammalian cells and introduced into a cell-free bacterial system (and vice versa) will produce the same proteins as in the original organism.

We should make the point that the universal nature of the genetic code does have some minor exceptions. For example, some stop codons can be different in prokaryotic and eukaryotic systems. In addition, cellular organelles (mitochondria and chloroplasts) have their own, separate, DNA which allows them to replicate. When the DNA of the mitochondria has been analysed, only a proportion of the tRNA molecules necessary for protein synthesis have been found. Thus mitochondria seem to use the same tRNA for a variety of amino acids—there is a very 'wobbly' coding system operating in these organelles. However, apart from some minor (but intriguing) exceptions such as these, all the available evidence indicates that the genetic code is universal among all the organisms that have been investigated. This, of course, is what is to be expected, if the range of modern species all have a common evolutionary origin.

This universality of code is not merely of abstract interest. As we shall see in Section 9, it has meant that some very ingenious experiments involving transfer of genetic information between organisms have been possible. These in turn have led to developments in genetic engineering of considerable practical importance.

## SUMMARY OF SECTION 7

1 On theoretical grounds alone, the minimum number of bases that could code for all of the amino acids is three—a triplet code.

2 Various experimental studies using synthetic RNA polymers have established that each amino acid is coded by a sequence of three bases in the mRNA. The triplet code (codon) for each amino acid has been determined.

3 Flexibility in base pairing between the tRNA and mRNA would allow for there being fewer distinct tRNA molecules than codons. One tRNA molecule may be able to pair up with more than one codon.

4 There are three specific stop codons which indicate when the polypeptide is to be released from the ribosome as a completed protein. There is also one codon which appears to initiate translation *as well as* coding for an amino acid (methionine).

5 Proteins may be modified once they have left the ribosome.

**SAQ 12** The synthesis of polypeptide Q is directed, at a ribosome, by mRNA molecule M.

(a) How many bases in M code for that part of Q comprising the N-terminal amino acid and the next nine adjacent amino acids?

(b) How many bases are there in that part of the coding strand of DNA that gives rise to the part of M described in (a)?

(c) Hydrolysis of Q shows that it contains 110 amino acids per molecule. Among these are found 18 of the possible 20 kinds of amino acid. What is the minimum number of different kinds of tRNA involved in the synthesis of Q?

**SAQ 13** Complete Table 7, using Table 6 to help you. (Remember the convention is that mRNA is translated from left to right.)

TABLE 7 For use with SAQ 13

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | TGT | | | | | non-coding strand } DNA double helix |
| | | | AGC | | | | coding strand } DNA double helix |
| | | | | | GUA | | transcribed mRNA |
| | UCC | | | GAG | | | tRNA anticodon |
| Met | | | | | | Trp | amino acid incorporated into the polypeptide |

**SAQ 14** Suppose evolution on Earth had occurred in such a way that (i) proteins consisted of 37 different kinds of amino acid, (ii) DNA contained six different kinds of base, and (iii) in all other respects DNA replication and protein synthesis occurred broadly in the ways described in Sections 5 and 6.

Under these circumstances what would be:

(a) the theoretical minimum number of bases per DNA codon;

(b) the theoretical minimum number of kinds of tRNA;

(c) the theoretical maximum number of kinds of tRNA (assuming three condons are stop codons and that codons consist of the minimum number of bases)?

**SAQ 15** Why would it have been unreasonable to invent a version of SAQ 14 in which *five* different kinds of DNA base were postulated?

# 8 MUTATIONS AND GENES

The notion that genes are much involved in the determination of phenotype has been a continuing theme of the biology Units in this Course. And, more specifically, the idea that genes can be altered by **mutation** has been important in considering possible mechanisms of evolution. From what you now know about the biochemical nature of the genetic material, can we attempt to interpret these concepts at the molecular level? Section 8.1 will address Question 5 posed in the Introduction, 'How does mutation change genes?'.

## 8.1 MUTATION

□ What is mutation?

■ Mutation is a heritable change brought about by an alteration in the genetic material of an organism (see Unit 19).

This definition implies that for the changes to be inherited, the mutation must occur in the cells that will go on to form gametes. In fact, mutations can also occur in somatic cells—but these will not have genetic consequences for the offspring of that organism and are not discussed here. Alterations in the genetic material occur in two main ways:

(i) chemical changes in the DNA;

(ii) radiation induced changes to the DNA.

Chemicals that cause mutation generally affect the DNA by altering or removing a single base. We shall not concern ourselves with the chemical details in this Unit, but concentrate on the effects.

Suppose part of a DNA molecule had the base sequence CGA CGG CTA CCA. The mRNA produced by transcription of this section of the DNA molecule would have the base sequence GCU GCC GAU GGU. By reference to Table 6, you can see that this mRNA would give rise to the following sequence of amino acids: alanine–alanine–aspartic acid–glycine. If a chemical caused the adenine in the first DNA codon to change to cytosine, i.e. CGA to CGC, do you think you would see a different amino acid at the left-hand end of the sequence? In fact, this DNA change would produce the mRNA codon GCG and, like GCU, GCG also codes for alanine. Thus although there has been a mutation of the DNA, in this case the mutation is not expressed. If however the cytosine in the first DNA codon were changed to guanine (i.e. CGA to GGA), the new mRNA codon would be CCU and this codes for a different amino acid—namely, proline.

Even a change as small as the alternation of a single base can have dire consequences.

□ Can you recall from Unit 22 an example of a drastic change in the function of a protein molecule caused by the replacement of one amino acid?

■ The abnormal haemoglobin found in people who suffer from sickle cell anaemia is discussed in both Unit 21 and Unit 22. The change in haemoglobin structure and function is caused by the replacement of glutamic acid in haemoglobin A by valine in haemoglobin S. (In fact, this is a change of one amino acid in haemoglobin A that takes place in each of two identical β chains, out of the four chains that constitute the haemoglobin molecule. See Plate 2 of Unit 21.)

Detailed analysis of haemoglobin taken from large numbers of people shows that there are hundreds of different kinds of single amino acid substitutions. The vast majority do not give rise to any observed phenotypic changes. This is because these amino acid changes have taken place in parts of the haemoglobin molecule that are not crucial for its function.

As well as causing the kind of base change already discussed, some chemi-

cals can cause mutations by bringing about the *deletion* of a base.

□ What effects would you predict from the deletion of one base in the DNA?

■ The codon sequence beyond the deletion ('to the right' of the deletion, so to speak) will be seriously affected. All the bases beyond this point will be out of phase, resulting in the whole amino acid sequence being changed beyond the position of the deletion.

The deletion—or for that matter, insertion—of one or two bases which alter the sense of the base sequence beyond the point of the base change, can be said to produce reading frame changes in the genetic message. A nice anology here would be to make sense of the present sentence if the printer had omitted the 'n' from 'analogy' but kept the word length the same, i.e. 'A nice aalogyh erew ouldb et om. . .' and so on. The change in the reading frame here, by the deletion of one letter, has made nonsense of the sentence.

For example, a sequence of DNA had the base sequence CGA CGG CTA CCA. The corresponding mRNA sequence is GCU GCC GAU GGU, leading to the following sequence of amino acids: alanine–alanine–aspartic acid–glycine. If, however, there was a deletion of a base, say the second cytosine, the DNA would now have the sequence CGA GGC TAC CA. Hence the corresponding mRNA would be GCU CCG AUG GU, and this codes for the amino acid sequence alanine–proline–methionine–.... The final amino acid, as you can see from Table 6, is likely to be valine, because the first two bases are GU.

ITQ 5 (a) What will be the sequence of amino acids coded by the following sequence of DNA bases: CCG TCT TTG CTC?

(b) In terms of amino acid sequence, what will be the result of a mutation that removes all the *purine* bases from the sequence of DNA in (a)? (Assume that the DNA will still be transcribed with these bases missing.)

The mutations brought about by radiation are rather different from those induced by chemicals. Such mutations take place when a DNA molecule interacts with ionizing radiation—radiation with sufficient energy to ionize matter—or with non-ionizing radiation such as u.v. In Unit 31, we shall consider in more detail the potential hazards of ionizing radiation—particularly the harmful products of nuclear reactions.

It is known that cells are able to counteract the effects of mutation to some extent by the use of specific enzymes. These enzymes bind to incorrect bases on one strand of the DNA and remove them. The undamaged complementary DNA strand is then used as a template to provide the correct base sequence. These **DNA repair mechanisms** even enable correct replication of the DNA to take place when X-rays have fused the two DNA strands together. When the strands are fused, DNA replication is impossible because the two strands cannot separate. What is thought to happen is that the normal enzymes necessary for DNA replication copy the DNA up to the point of fusion. They then skip to some point beyond the fused region, which may be hundreds of bases along the DNA, and start the copying process again. This leaves a large section of DNA uncopied. Repair enzymes then 'unwind' the uncopied DNA, breaking the chemical bonds holding the fused strands together, and make complementary copies which form into new double helices. Other enzymes then 'stitch' these helices into the new DNA.

These repair systems are complex and you do not need to understand how they may work. What you should be aware of, though, is that errors can be introduced into the DNA as a result of mutation, and that mechanisms exist for the cell to repair this genetic damage. These repairs are not perfect, which is why a pool of new genetic information is built up in the DNA. It is this changed genetic information which is the starting point for evolution.

GENE

STRUCTURAL GENES

## 8.2 WHAT IS A GENE?

You may recall from Unit 20 that the concept of a **gene** is linked in some way with the physical structures known as chromosomes—genes are said to be linked if they are found on the same chromosome. It should be clear from this Unit that we could consider a gene as being a length of DNA that carries the code for a particular polypeptide chain; and, indeed, there are very many genes of this kind. They are often termed **structural genes** because they contain the code for a protein *structure*.

But there are other kinds of gene, too. We know that DNA carries the code for both tRNA and rRNA as well as mRNA, so there must also be genes for these other classes of ribonucleic acid. In addition, there are the crucially important lengths of DNA that control whether other DNA sequences are or are not transcribed—these are the control genes you met in Section 4. The common characteristic in this range of different kinds of gene—those that make mRNA, rRNA or tRNA, and those that control transcription—is that they all code for specific molecules which are involved in the production of proteins. Taking this as a definition of a gene, can we start to explain dominant and recessive characters (Unit 20) in molecular terms?

A simple example here would be eye colour in *Drosophila*. The eye colour is normally bright red, and this is a dominant character associated with one gene, *E*. There are also mutant forms which have cinnabar coloured eyes. Genetic studies show that the flies with such eyes are homozygous recessive with respect to the eye colour gene *e*. The red colour is due to a red pigment produced by a series of reactions, one of which involves a particular enzyme and a red-pigment precursor. As you know, enzymes are proteins, and thus are coded for by a sequence of DNA—a gene. In the mutant form of *Drosophila*, an altered DNA sequence gives rise to an altered enzyme, which gives rise to the different eye colour. Thus there is a link between the gene at the molecular level and a gene defined by its phenotypic consequences. This is straightforward, but how can we explain the phenotype of the heterozygous flies—*E e*? In these flies, both a normal metabolic pathway and an abnormal one will, in theory, be present.

In this case, both red *and* cinnabar pigments are being produced. Detailed chemical analysis shows reduced amounts of these two pigments compared with the two homozygous conditions. Even though both pigments are present in the heterozygous situation, the red pigment masks the cinnabar pigment so the eye appears red. You can think of this in terms of the effect you would see if you painted a pink surface with red paint: despite the fact that the pink paint is still there, all you would see is the red paint. We say that the red phenotype is *dominant* to the cinnabar phenotype. This use of the term dominant is more correct than saying that the genes themselves are either dominant or recessive. It is not the gene but the gene *product* which is dominant or recessive in the heterozygous state.

Another example of how we can explain dominance in molecular terms is in the round and wrinkled pea seeds you met in Unit 20 when considering Mendel's observations. Whether seeds are round or wrinkled is determined by a single gene, and the character of 'roundness' is dominant to that of 'wrinkled'. The reason why seeds are round or wrinkled is related to the concentration of sugar in the seed. We need not go into detail, but if seeds contain more than a certain level of sugar they will absorb water during their formation, swell and thus produce a smooth seed coat—and hence round seeds. If the sugar is below this level, the developing seeds do not absorb enough water and the seed coat takes on a wrinkled appearance.

The formation of sugar involves a metabolic pathway with a number of enzymes. If one enzyme in this pathway is altered, a reduced level of sugar is made in the seed. Biochemical analysis of the round and wrinkled seeds shows that the heterozygous round seeds have only half as much of this enzyme as the homozygous round seeds, while the wrinkled seeds have virtually undetectable levels of an active form of this enzyme. Thus by measuring the level of this enzyme, we can tell if the seeds are heterozygous, homozygous dominant, or homozygous recessive.

□ Why do you think the heterozygous round seeds have only half the amount of the enzyme involved in the metabolism of sugar, compared with the homozygous round seeds, and why do the homozygous recessive seeds have virtually none?

■ The heterozygous round seeds have only one gene for this enzyme, whereas the homozygous round seeds have two genes that carry the code for this enzyme. Since there are two genes in the round homozygote, twice as much of the enzyme will be produced as in the heterozygote with only one gene. The wrinkled seeds (homozygous recessive) have no genes coding for this enzyme, and hence no active enzyme and virtually no sugar.

Although the heterozygotes have only half the level of enzyme this is still adequate to allow the synthesis of sufficient sugar to 'create' the round-seed phenotype.

The red/cinnabar eye colour in *Drosophila* and the round/wrinkled seeds in the pea are two simple examples of how we can start to explain the dominance of genes in molecular terms. You should note, though, that in many cases the picture is more complicated. There can be a mixing of gene products to produce heterozygotes which have a character intermediate between that seen in the dominant and recessive situations. In addition, what we observe as phenotypic differences are often due to the interaction of a large number of different genes and hence gene products. These complex kinds of inheritance, though undoubtedly of enormous biological significance, were not dealt with in Unit 20 and will not be considered further in this Unit.

## SUMMARY OF SECTION 8

1 Mutations brought about by a chemical agent may involve either a change in or the deletion of a single base, or the loss of a sequence of DNA bases. Mutation is also brought about by non-ionizing radiation such as u.v., and by ionizing radiation (which is discussed in Unit 31). Mutation in the cells which will form gametes generally leads to heritable changes in the DNA of the offspring of an organism. Some repair of the DNA to correct errors introduced by mutation is possible.

2 The term gene is difficult to define precisely in molecular terms, though a useful definition would be to say that a gene is a length of DNA that codes for molecules involved in the production of proteins.

**SAQ 16** If a mutation occurred in a sequence of DNA bases coding for a protein, so that . . . AAA GAG . . . was changed to . . . TAA GAG . . . , what change would you expect to see in the amino acids that are incorporated in the polypeptide at this point?

**SAQ 17** The enzyme ribonuclease, which breaks down RNA molecules, has a sequence of 124 amino acids. For the purposes of this question you can assume that the active site of this enzyme is determined specifically by the sequence of the last 110 amino acids. The DNA sequence in the 6th, 7th and 8th codons is

| CGA | TTC | AAA |
|---|---|---|
| 6th codon | 7th codon | 8th codon |

Explain what is likely to occur, with respect to the function of the enzyme, if the following mutations occur in the DNA coding for the sixth, seventh and eighth amino acids:

(a) The first base of the sixth codon in the sequence is replaced by guanine.

(b) The first base of the seventh codon is changed to adenine.

(c) The last base of the eighth codon is changed to guanine.

(d) The middle base and last base of the seventh codon are deleted.

(e) All three bases of the eighth codon are deleted.

**SAQ 18** Unit 20 describes a colour character seen sometimes in maize cobs, where some seeds are purple but others are white. The white colour is recessive to purple. Explain how a gene might act to produce the white colour.

# 9 DNA—A CONTEMPORARY VIEW

Previous Sections have described the basic features of DNA structure and protein synthesis. In this Section we see how current knowledge has progressed dramatically from the ideas of the 1960s. Material in Section 9.2 is relevant to the TV programme and forms a set of notes for it.

## 9.1 RELATIONSHIP BETWEEN DNA, mRNA AND PROTEIN

It was assumed for many years that the sequence of codons in the DNA of a gene coding for a given protein was in the form of *one continuous thread.* This view was shattered in 1977 when investigators in several different laboratories found that, for some genes from eukaryotic organisms, this was not so.

As you can see from Figure 38a, a gene often consists of **exons** interspersed with **introns**. *Ex*ons get their name because their base sequence is ultimately *ex*pressed in protein production, while the *inter*vening bits, the *intr*ons, do not code for anything. Discontinuous genes of this type are often referred to as **split genes**. (We should make it clear that such split genes have not been found in prokaryotic cells.)

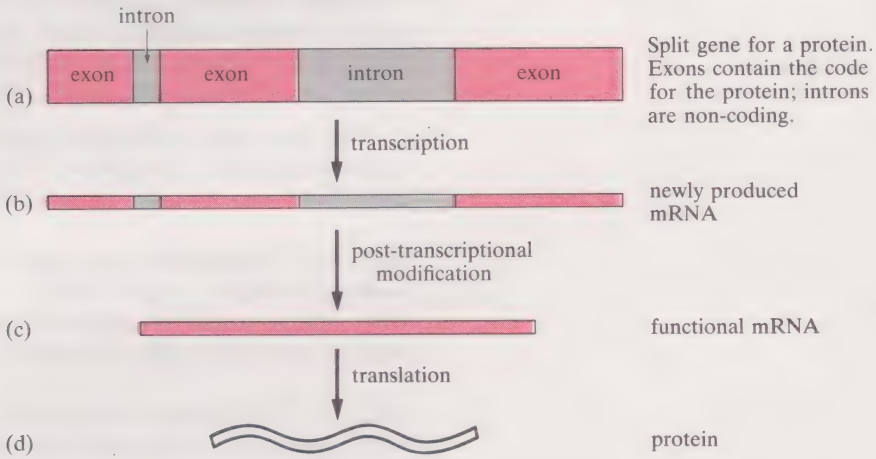



FIGURE 38 The transcription and subsequent translation of a split gene. Note that both exons and introns are initially transcribed. The newly produced mRNA is then modified to the form in which it is translated.

The mRNA that is first produced by transcription from such a gene is found to be much longer than the true mRNA that is ultimately translated.

☐ What must happen, therefore, to mRNA newly transcribed from a split gene before it is translated?

■ It must be modified and shortened in some way first.

This process of 'tailoring out' the unwanted parts of the newly transcribed RNA and then 'splicing' together the required pieces that remain is called **post-transcriptional modification**. The colloquial phrase **tailoring and splicing** is frequently used to describe the enzyme controlled processes involved in

this modification. Once modified in this way, the mRNA proceeds to the translation stage. Figures 38b to 38d show the modification and subsequent translation of RNA transcribed from a split gene.

A good example of the complexity of split genes is the gene which codes for egg white protein, ovalbumin. This has *eight* exons with *seven* intervening intron sequences!

Thus a simplistic picture of DNA coding directly for a protein is rather misleading. To further complicate the story it is now known that, as well as intron sequences *within* genes, maybe as much as 80% of the 'non-gene' DNA has no known coding function at all! This is illustrated in Figure 39. This Figure is purely diagrammatic—estimates of the proportion of DNA that carries no code, and of the relative proportions of DNA that code for RNA (**coding DNA**), show great variation in different species.

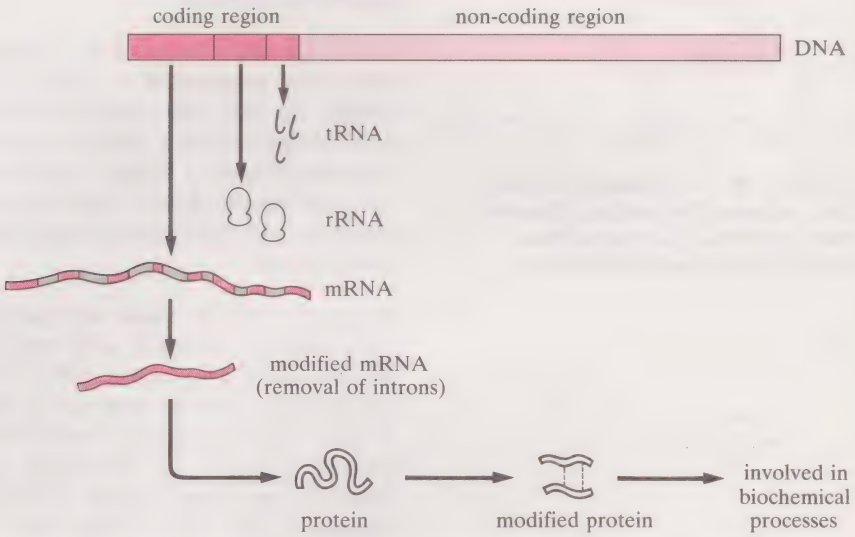



FIGURE 39 The relative proportions of DNA which are coding and non-coding, and the products of the coding regions of the DNA.

This Unit is not the place to discuss the various theories that attempt to account for this huge amount of **non-coding DNA** in living organisms. However, these non-coding regions have the most remarkable characteristic of being highly variable in base sequence from person to person. So much so, in fact, that the variability forms the basis of what is now called **genetic fingerprinting**. By comparing the base sequences in the DNA of (say) an old blood stain with that in the DNA of a suspect person, forensic scientists make use of the uniqueness of these hypervariable non-coding regions to establish whether the blood of the stain is or is not that of the person in question, with a certainty of 4 000 000 to 1. In like manner (bearing in mind that 50% of a child's genes come uniquely from each parent), it is plainly possible to say in a way which is effectively indisputable (30 000 million to 1!) whether a particular person is or is not the child's parent.

Another application of the modern and detailed knowledge of protein synthesis that this Section has touched on, is concerned with the production of quantities of proteins important to humans, by techniques called **genetic engineering**. Insulin manufacture provides an excellent and important example—and warrants a Section of its own (Section 9.2).

## 9.2 BACTERIA—FACTORIES FOR HUMAN PROTEINS (TV PROGRAMME)

In recent years, there have been major developments in what has become known as genetic engineering. In outline, this involves transferring genetic information between organisms. If a gene for a particular protein in a eukaryotic organism can be transferred into a prokaryotic organism (which

PLASMIDS

grows and divides rapidly), it is possible to 'manufacture' large quantities of this protein. Because the genetic code is to all intents and purposes *universal*, the prokaryotic cell (usually a bacterium) will simply transcribe the new genetic information to mRNA, which is then translated to make the new protein.

Scientists have also developed techniques that enable them to introduce genetic material into isolated plant cells. These cells then grow and develop into adult plants in which *all* the cells contain the new genetic information. Thus, it has been possible to transfer genes which confer resistance against disease from one species of plant to another.

Genetic engineering has been used to make a wide range of biological molecules, for example hormones and the anti-viral agent interferon. Work is currently progressing on the production of safer vaccines using genetically engineered bacteria.

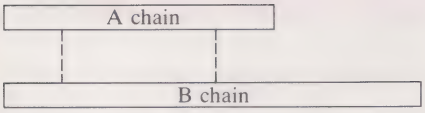



FIGURE 40 A schematic diagram of the structure of the functionally-active insulin molecule. The dashed lines denote the two disulphide bridges.

One of the main topics of the TV programme 'DNA' concerns how bacteria have been 'engineered' to make a protein not found in bacteria—human insulin. As you may remember from Unit 22, the functional insulin molecule consists of two polypeptide chains, A and B, joined together by disulphide bridges: a simple schematic representation is shown in Figure 40. You also known from Unit 23 that insulin is a hormone produced by the pancreas, and it is crucially involved in the regulation of glucose concentration in blood.

It appears that, in some populations, as many as five in every 100 people have defects associated with insulin function; this leads to one of several different kinds of diabetes. Often, in these cases, insulin is produced in much reduced quantities or even not at all; if untreated, this condition would lead to very high blood sugar levels (the hyperglycaemia you met in Unit 23) with potentially very damaging consequences. Treatment in all but the mildest cases depends upon regular and life-long injections of a solution of insulin. In the past, insulin has been derived from animal tissues (pig pancreas). This can have side effects, as pig insulin has a slightly different amino acid sequence to human insulin. With a disease as common as diabetes, requiring medication on the scale it does, it is not surprising that the idea of putting the human insulin gene into bacteria, and so manufacturing the pure human hormone in whatever quantities are required, has been an alluring and potentially profitable goal. And, indeed, success has been achieved—a succession of developmental stages in the early '70s and clinical trials in 1980 led to the launch of a commercial product in July 1983.

How was it done? The TV programme provides some of the details, which we set out again here. The desired product—functionally active insulin—is the two-chained molecule shown in Figure 40; the gene that codes for it is a split gene.

☐ Recall what this means, expressing your answer in terms of the protein insulin.

■ The insulin gene contains exons that are ultimately expressed as protein, and introns that are non-coding.

In fact, the gene for insulin has two introns and two exons. However, the situation is just a little more complicated than you might have been expecting from the two-chained structure of Figure 40. Look at Figure 41 and you will see why!

Here, (iv), (v) and (vi) are all proteins; the last of these is functional insulin. When functional insulin is secreted from storage granules inside the pancreatic cells, it is formed from a precursor molecule called *pro-insulin*. This, in turn, is formed from an even longer protein with the even longer name of *pre-pro-insulin*. It is pre-pro-insulin (shown in Figure 41(iv)) that is translated from the functional mRNA (shown in Figure 41(iii)). This, of course, is formed by post-transcriptional modification of the mRNA newly transcribed from the split gene (shown in Figure 41(ii)).

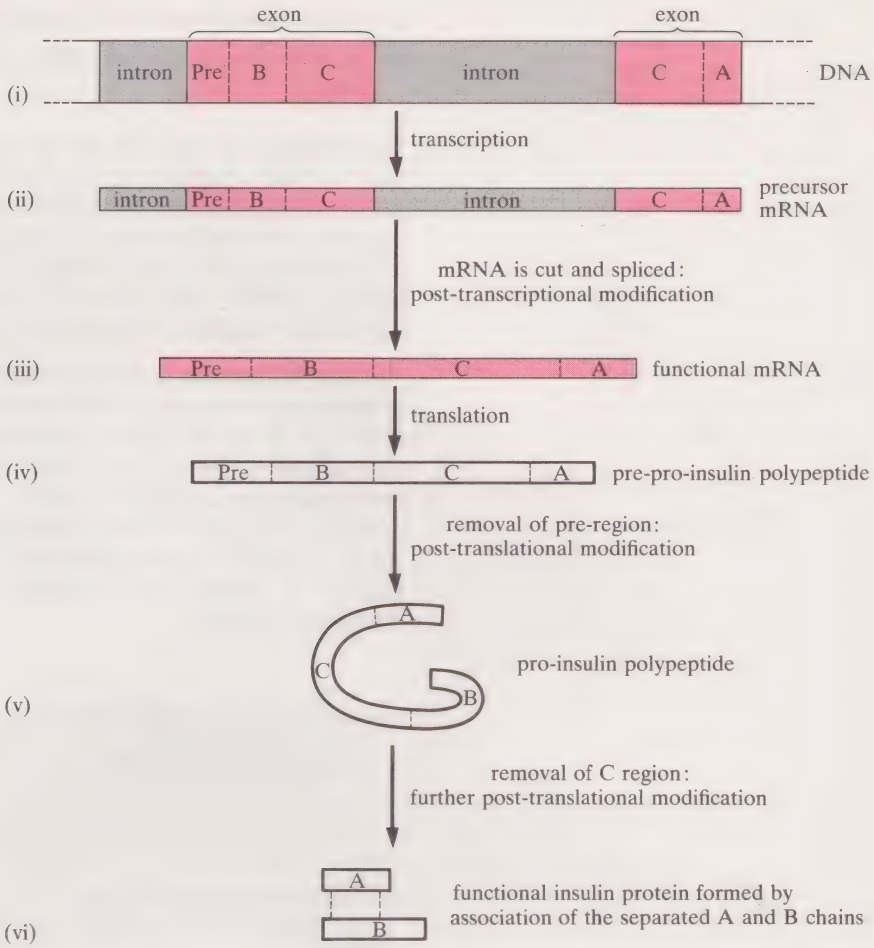



FIGURE 41 From gene to insulin.

Bacteria are perfect chemical factories. They multiply rapidly, they are cheap to maintain and—very important—unlike eukaryotic cells they have DNA which is not associated with protein molecules in the complex way described in the chromatin story of Unit 20. Because the bacterial DNA is 'naked', added DNA of the desired type can readily be incorporated in some part of the genome of the bacterium. Thus, in essence, all that genetic engineering involves is introducing into the bacterial DNA a sequence of DNA that codes for the protein required.

As you will see in the TV programme, this simple theory is difficult to put into practice! It is in fact quite easy to get bacteria to 'pick up' sequences of DNA. Bacteria contain small circles of DNA known as **plasmids**. These replicate in phase with the bacteria and carry genes for a variety of proteins. The plasmids can be extracted, broken open and mixed with new DNA. Under the right conditions, the new DNA is incorporated into the plasmid, which can then be taken up by other bacterial cells. These plasmids will then be translated giving, along with bacterial protein, the appropriate human mRNA and human protein.

□ What problem would there be if bacteria were allowed to make direct use of the entire human insulin gene?

■ The DNA sequence for insulin contains introns. Thus a non-functional, totally unnatural protein would be made.

One way around this might be to isolate from the cytoplasm the functional mRNA that codes for insulin; this will not contain transcripts of the introns. Once extracted (by ways not described here), transcription can be worked 'in reverse' to give DNA of the type from which it would have come. This would, of course, be the original insulin gene minus the introns. This copy DNA, as it is called, could be introduced into plasmids and so into bacteria. This would then be transcribed and translated into protein.

☐ This technique would be no use either. Why not?

■ Because the protein produced would be pre-pro-insulin. This is not biologically active.

The solution, as you will see in the TV programme, depended on getting back to first principles to develop a theory, plus a good deal of sophisticated experimentation. Because we know the structure of biologically active insulin in terms of its amino acid sequence, it should be possible to make by *artificial means* the DNA coding for each of the two strands in a functional insulin molecule. This synthetic molecule is the one that should (and did) work when supplied to the plasmids.

The overall process is summarized in Figure 42. The most obvious feature is that separate plasmids and entirely separate bacterial cultures are used to make the A and B chains of insulin. Stages (e) and (f) indicate a problem that we have chosen not to dwell on in the preceding paragraphs: at translation, unwanted *bacterial* protein (transcribed from the plasmid DNA) is produced, attached to the insulin (A or B) chain protein. This has to be got rid of. Once this has been achieved, the separately produced A and B chains have to be joined into the active molecule by the formation of the two disulphide bridges.

(a) Manufacture of synthetic sequences of bases for the A and B polypeptide chains of insulin.

(b) Incorporation of base sequence into plasmid next to a specific bacterial gene.

(c) Introduction of plasmid into *E.coli*. The *E.coli* cells are then allowed to grow.

(d) Bacterial gene transcribed along with insulin amino acid sequence.

(e) Separation of bacterial protein from insulin protein.

(f) Joining of A and B chains to form a functional insulin molecule.

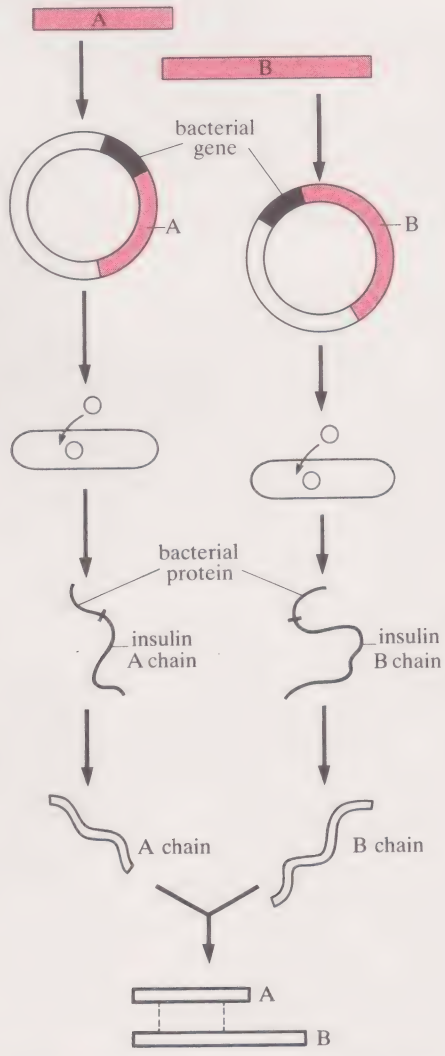



FIGURE 42 The stages involved in an early method of the genetic engineering of bacteria to make human insulin.

We should make it clear that this is only one solution to producing insulin and, although it was used initially for commercial production, the present technique uses bacteria which have been engineered to transcribe and translate the DNA sequence for pro-insulin. After extraction of this molecule from the bacteria, specific enzymes are used to remove the 'C' region.

As regards genetic engineering, the aim of the programme is to show you that:

(a) it is possible to make DNA sequences for specific proteins, once the amino acid sequence is known;

(b) bacteria can be used as hosts for this DNA, and can be induced to translate it to make mRNA and then proteins;

(c) the decoding of artificial DNA by bacteria is further evidence of the almost universal nature of the genetic code.

## SUMMARY OF SECTION 9

1 Over 80% of the DNA of most eukaryotic organisms consists of non-coding sequences of DNA that are not associated with genes and have no known function.

2 In eukaryotes, the DNA of most genes contains non-coding base sequences. These non-coding lengths within a gene are called introns, while the coding lengths are called exons.

3 Newly synthesized nuclear mRNA is longer than is needed to code for a specific protein, because both exons *and* introns are transcribed. Post-transcriptional modification (involving tailoring and splicing) converts this initial mRNA into functional mRNA. Normally, only the latter is translated.

4 Many proteins, such as insulin and ovalbumin, are modified after translation.

5 The genetic code (with the exception of DNA in cellular organelles) is universal. This is clearly shown by the fact that functional genetic material can be transferred between different species.

6 It is possible to make artificial DNA and to introduce this into bacteria, and thus get the bacteria to make mammalian proteins.

**SAQ 19** Explain why each of the following statements is incorrect:

(a) If you can determine the amino acid sequence of a protein, it is always possible to work out the base sequence in the DNA which codes for this protein.

(b) A length of mRNA for a protein, extracted from the nucleus of a mammalian cell, would have approximately three times the number of bases as the number of amino acids in that protein.

(c) A change in the nature of even one base in the DNA will always alter the sequence of bases in the functional mRNA molecule produced from that DNA.

(d) Bacteria can be made to make human proteins by altering the base sequence of existing bacterial genes.

(e) In a cell-free system, translation of the mRNA for pre-pro-insulin will give rise to a functional insulin molecule.

**SAQ 20** Write a short set of notes which explain how bacteria were initially used to manufacture human insulin.

# 10 CONCLUSION

Stepping back from the detail of the Unit, have the questions asked in Section 1 been adequately answered? The questions (and the Sections involved) were as follows:

1 What is the chemical nature of a gene? (*Sections 2 and 3*)
2 Why are cells different from each other? (*Section 4*)
3 How do genes replicate? (*Sections 3 and 5*)
4 How does genotype influence phenotype? (*Sections 6 and 7*)
5 How does mutation change genes? (*Section 8*)

You should, by now, have a reasonable view of the basic answers that some forty years of research in molecular biology have provided. Yet the answers given in this Unit are incomplete—partly because only a fraction of what is known can be included in one Unit, and partly because this area of biology is still at the frontiers of current research: treatment of AIDS, cancer therapy, hormone production through genetic engineering, and numerous other areas. There is a whole range of fundamental questions with answers that are still fragmentary: the switching on and off of genes during development from zygote to adult is one example, as is a possibly related one—the matter of ageing and its relationship to the molecular biology of the gene.

But what of the questions we raised in Section 1? A lengthy summary at the end of a long Unit is not perhaps appropriate, but a few broad sweeps of the brush may be useful.

*Question 1* You should have been convinced by the evidence that DNA is the genetic substance, and the picture you have of the chemical structure of DNA (and of RNA) should be clear enough.

*Question 2* This question has, perhaps, had the least satisfactory answer, partly because what is known about the 'switching' mechanism is complex, and partly because much is still unknown. This should have been compensated for, however, by the extraordinary import of Gurdon's toad experiments: for any individual, *all cells have all genes*, it appears—yet at just the right times only the necessary few are switched on.

*Question 3* Meselson and Stahl's famous experiment has provided a framework for answering the question. Their experiment demonstrates that, in replication, each strand of DNA acts as a template for the formation of another strand, through complementary base pairing. As a point of major importance, you should have seen how well the structure of the DNA double helix matches its function as a molecule that can be replicated.

*Question 4* This question is, of course, the molecular-biological keystone of genetics and biochemistry. The DNA of 'switched-on' genes manifests its genetic message (that is, the encoded information) by calling specific proteins into being, through the agency of transcribed RNA and the whole complicated ribosomal paraphernalia of translation. These proteins then 'do' the biochemistry that shapes the phenotypes we observe as inherited characteristics. Once again, the structure of DNA can be seen to match, most elegantly, its biological role as a 'transcribable information carrier'.

*Question 5* This question is concerned with the mutability of DNA. It is not only relevant to modern concerns about the unwanted consequences of radiation damage, but also to the fundamental phenomenon of the 'creation of variation' that underlies all evolution. And once again, the structure of DNA—partly very stable yet partly not—matches its role as a molecule that is largely but not quite unchanging.

Whatever were the first forms of life some 4000 million years ago, it is certain that they were very simple. Increasing complexity has been a characteristic of evolutionary change ever since. In all organisms, including the very large number of now extinct species, the diversity of phenotypes has arisen through the molecular-biological mechanisms discussed in these pages. The result is the several million species now living. The way they interact, competitively or cooperatively, is a major concern of the science of *ecology*—the subject of Unit 25.

# OBJECTIVES FOR UNIT 24

After you have worked through this Unit, you should be able to:

1 Explain the meaning of, and use correctly, all the terms flagged in the text.

2 Give brief accounts of two lines of evidence which support the view that DNA is the carrier of genetic information: (i) Hershey and Chase's experiment with the $T_2$ virus and *E. coli*; (ii) measurement of the mass of DNA in the nucleus of a gamete or a somatic cell. (*SAQs 1 and 2*)

3 Give a brief account of the structure of DNA. (*SAQs 3 and 4*)

4 Do simple calculations to estimate the relative proportions of the different bases in DNA. (*SAQs 3 and 4*)

5 Give a brief explanation of the phases of the cell cycle. (*SAQ 5*)

6 Explain how Gurdon's experiments on transferring nuclei from one cell to another, support the hypothesis that there is no loss of genetic information as cells become differentiated. (*SAQ 6*)

7 Explain briefly how DNA replicates. (*ITQ 1*; *SAQ 7*)

8 Outline the experimental evidence that confirms the semi-conservative replication of DNA. (*ITQ 2*; *SAQ 8*)

9 Make predictions about the results of extending the Meselson and Stahl experiment beyond generation 1. (*ITQ 2*; *SAQ 8*)

10 Give a brief account, including diagrams, of the structure of RNA, and show how this differs from that of DNA. (*SAQ 9*)

11 Outline the roles of DNA, mRNA, tRNA and ribosomes in the process of protein synthesis. (*ITQs 3 and 4; SAQs 10, 11 and 12*)

12 State, in their correct sequence, the principal events occurring during the synthesis of a protein. (*ITQs 3 and 4*; *SAQ 11*)

13 Knowing the sequence of DNA bases coding for a protein, give the corresponding sequence of mRNA bases. (*ITQ 5*; *SAQ 13*)

14 Given either a DNA or an mRNA codon, deduce the possible tRNA anticodon. (*SAQ 13*)

15 Translate an mRNA sequence into a sequence of amino acids, given the relevant information about the genetic code. (*ITQ 5*; *SAQ 13*)

16 Do simple calculations showing the relationship between the number of bases in a codon and the number of possible codons. (*SAQs 14 and 15*)

17 Explain how mutation can change the nature of DNA bases. (*ITQ 5*; *SAQs 16 and 17*)

18 Explain at the molecular level what is meant by a gene. (*SAQ 18*)

19 Explain briefly how mRNA is modified after transcription and give an example of how a protein can be modified after translation. (*SAQ 19*)

20 Outline the key steps involved in getting bacteria to manufacture human insulin. (*SAQ 20*)

# FURTHER READING

A good basic text which covers most of the topics in this Unit is one in the Institute of Biology, Studies in Biology Series No. 83:

Clark, B. F. C. (1984) *The Genetic Code and Protein Biosynthesis*, Institute of Biology and Edward Arnold.

For further reading on the topic of the cell cycle and cell growth you could read another Studies in Biology booklet No. 148:

Wheatley, D. N. (1982) *Cell Growth and Division*, Institute of Biology and Edward Arnold.

For a detailed account of aspects of genetic engineering you could look at the following book—though it is written at a much higher level than the material in Section 9 of this Unit:

Watson, J. D., Tooze, J. and Kurtz, D. T. (1983) *Recombinant DNA—A Short Course*, Scientific American Books, W. H. Freeman & Co.

The following text book provides a full coverage of the molecular biology of DNA. If you are fascinated by this topic this is well worth looking at.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. (1987) *Molecular Biology of the Gene*, Vol. 1, Benjamin/Cummings.

Finally, a great book which gives an exciting account of the discovery of the structure of DNA is:

Watson, J. D. (1968) *The Double Helix*, Weidenfield and Nicholson.

# ITQ ANSWERS AND COMMENTS

**ITQ 1** You might have suggested the following:

(a) The strands of the double helix can separate, because the two strands are only held together by the weak hydrogen bonds between the bases.

(b) The free nucleotides in the surrounding solution could line up against the exposed bases along each of the separated strands, in a manner dictated by the base-pairing rules (A–T and C–G). By this means, each old (and now separated) strand would act as a template for the formation of a new strand complementary to it. The nucleotides aligned against each old DNA strand would link together, by means of covalent bonds between the phosphate and sugar groups. Thus two new double helices would have formed, identical to each other and to the original double helix.

If this is what you did suggest, then your suggestion exactly matches current theory!

**ITQ 2** Generation 1 will have helices with one 'heavy' and one 'light' strand. When these divide to give generation 2, the daughter helices will be of two types. One will have an original heavy strand and a new light strand, the other will have two light strands (one new, one from generation 1).

**ITQ 3** The sequence of bases in mRNA formed by transcription is the complement of the sequence of bases in the DNA strand used as a template. The normal base-pairing rules ensure this complementarity. In mRNA, uracil takes the place of thymine.

**ITQ 4** Your completed paragraphs should have incorporated the words in italics.

DNA carries a code for *mRNA*. After a process known as *transcription*, these molecules move into the *cytoplasm* of the cell via pores in the nuclear membrane. In the cytoplasm, they act as templates for the production of *polypeptide* molecules by the process of *translation*.

A sequence of three *bases* in mRNA is called an mRNA *codon*. The sequence of these *codons* [or *bases*] along a particular mRNA molecule determines the specific sequence of *amino acids* in the *polypeptide* coded for by the *DNA* of the gene and its complementary *mRNA*.

Next, in a progressive manner, these *amino acids* [or *tRNA molecules*] position themselves over the appropriate mRNA codons. At each step, through the catalytic action of specific *enzymes* and with conversion of *ATP* to *ADP* and $P_i$, an amino acid is added to the growing *polypeptide* [or *peptide*] chain.

Ultimately, when the *stop* codon is reached, the completed *polypeptide* [or *protein*] molecule is released into the cytoplasm. Several *ribosomes* are able to move along one mRNA molecule, each bearing a progressively longer *peptide* chain. The overall structure—consisting of an mRNA molecule, several *ribosomes*, and *peptide* chains of different lengths—is called a *polysome*.

**ITQ 5** (a) Glycine–arginine–asparagine–glutamic acid.

If the sequence of DNA is CCG TCT TTG CTC, this will give the mRNA sequence:

GGC AGA AAC GAG

(Remember the base-pairing rules, C–G and A–T or U.) Referring back to Table 6, this mRNA sequence codes for the amino acid sequence:

glycine–arginine–asparagine–glutamic acid

(b) Although the amino acid sequence would still start with glycine, the remainder of the sequence would be very different.

The purines are the bases A and G (see Section 3.2). Therefore if the purines are removed from the DNA, bases A and G are deleted. In the DNA sequence in (a), we therefore 'cross out' each G (since there are no As in this sequence). Thus, the DNA sequence becomes CCT CTT TCT C after the mutation, which gives the mRNA sequence GGA GAA AGA G. This will code for the following sequence of amino acids:

glycine–glutamic acid–arginine ...

You can see that the removal of just two bases has drastically altered the sequence of amino acids. Although the sequence still starts with glycine, glutamic acid and arginine are now in different positions relative to each other and asparagine is no longer present.

Note that this hypothetical question describes a very unlikely situation—a mutation which removed *all* purine bases would, in practice, result in a section of DNA that could no longer be transcribed because too many bases would be missing. It would, however, be possible for an mRNA molecule to be made if just a small number of bases were deleted from the DNA molecule.

# SAQ ANSWERS AND COMMENTS

**SAQ 1** Substantial amounts of protein would enter the bacterial cells, while the nucleic acid would remain outside them.

**SAQ 2** (a) 5.4 picograms.

Gametes have half the number of genes present in somatic cells (one from each gene pair). Hence the nucleus of a gamete has half the mass of DNA present in the nucleus of a somatic cell of the same species.

(b) (i) 2.7 picograms, (ii) 2.7 picograms, (iii) 170 picograms.

The reasoning in (i) and (ii) is identical with that in SAQ 2(a).

(iii) Each somatic cell of the turtle will contain the same amount of DNA. Since the cells lining the intestine each contain 5.3 picograms, it follows that a group of 32 cells will have $32 \times 5.3$ picograms $\approx$ 170 picograms of DNA.

**SAQ 3** (a) (i) 1 : 1; (ii) 1 : 1; (iii) 1 : 1.

These ratios can be determined by counting the bases in the piece of double helix in Figure 12, i.e. 6 A : 6 T (1 : 1), 7 G : 7 C (1 : 1), 13 pu : 13 py (1 : 1). You can also deduce the answer, without counting the various bases, from the base-pairing rules: A–T and C–G. There will be same amount of A as T and of C as G; this means that there will be the same amount of (A + G) as of (C + T), and hence equal amounts of purines and pyrimidines.

(b) You can predict that each of these three ratios will again be 1 : 1. *Any* segment (or entire molecule) of DNA double helix has A : T = 1 : 1, G : C = 1 : 1 and pu : py = 1 : 1, because of the base-pairing rules.

**SAQ 4** (a) 160; (b) 57; (c) 57; (d) 23; (e) 183; (f) 160; (g) 160; (h) 160; (i) 80.

(a) As there are 80 purine bases, there must be 80 pyrimidine bases, i.e. 160 bases altogether.

(b) 23 of the pyrimidine bases are cytosines, therefore the remaining (80 − 23) = 57 bases are thymine.

(c) and (d) By the base-pairing rules, A = T and G = C, so there are 57 adenines and 23 guanines. As a check, note that 23 + 23 + 57 + 57 = 160.

(e) There are always two hydrogen bonds between A and T, and there are 57 A–T pairs in fragment Y, so there are 114 hydrogen bonds involved in the A–T pairs. There are always three hydrogen bonds between C and G, and there are 23 G–C pairs in fragment Y, so there are 69 hydrogen bonds involved in the G–C pairs. Thus, there is a total of 114 + 69 = 183 hydrogen bonds.

(f), (g) and (h) As there are 160 bases, there must be 160 phosphate groups, 160 nucleotides and 160 deoxyribose molecules.

(i) As there are 160 bases, there must be half that number (80) of complementary base pairs.

**SAQ 5** Your diagram should be similar to Figure 13. The main events are cell growth, DNA replication, a phase during which cellular components are organized prior to division, and mitosis (cell division).

**SAQ 6** (a) False. Although this supposed result would have failed to show that all genes are present, it would not have given positive support to the contrary hypothesis. This is because there are other perfectly feasible explanations for the negative result; for example, the experiments might have been carried out under unsuitable conditions. Categorical *positive* statements are frequently suspect in biology. Always bear this in mind when looking at assessment material associated with biology!

(b) True—in this hypothetical example! If many different kinds of somatic cell from the same organism had different weights of DNA per nucleus, this would support the hypothesis that differentiation is the result of a differential loss of genes. (Again, it would not *prove* it: other explanations might apply that would still make the contrary hypothesis tenable.) Remember though that loss of genetic material has not (so far) been found to happen as differentiation occurs.

(c) True. Gurdon's work supports the view that each somatic cell in an organism has all the chromosomes, and therefore the cells of the human big toe almost certainly have the same genes as the cells of the human retina.

**SAQ 7** Figure 43 is the completed form of Figure 20, and Table 8 is the completed form of Table 2. Bases in black are those given in the question; bases in red are those that can be deduced by the base-pairing rules. Note that nothing can be deduced about the bases in positions 1 and 2.

TABLE 8 For use with the answer to SAQ 7

| Position | Identity | Position | Identity |
|---|---|---|---|
| 1 | not known | 2 | not known |
| 3 | adenine | 4 | thymine |
| 5 | thymine | 6 | adenine |
| 7 | guanine | 8 | cytosine |
| 9 | guanine | 10 | cytosine |
| 11 | cytosine | 12 | guanine |
| 13 | thymine | 14 | adenine |
| 15 | cytosine | 16 | guanine |
| 17 | cytosine | 18 | guanine |
| 19 | thymine | 20 | adenine |
| 21 | cytosine | 22 | guanine |

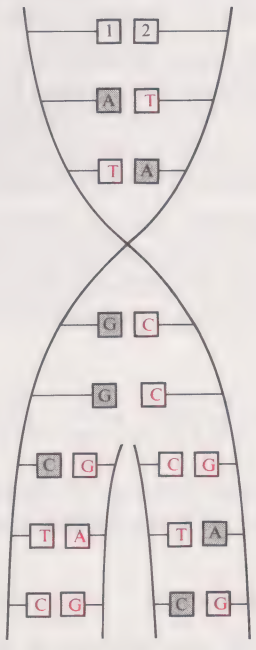
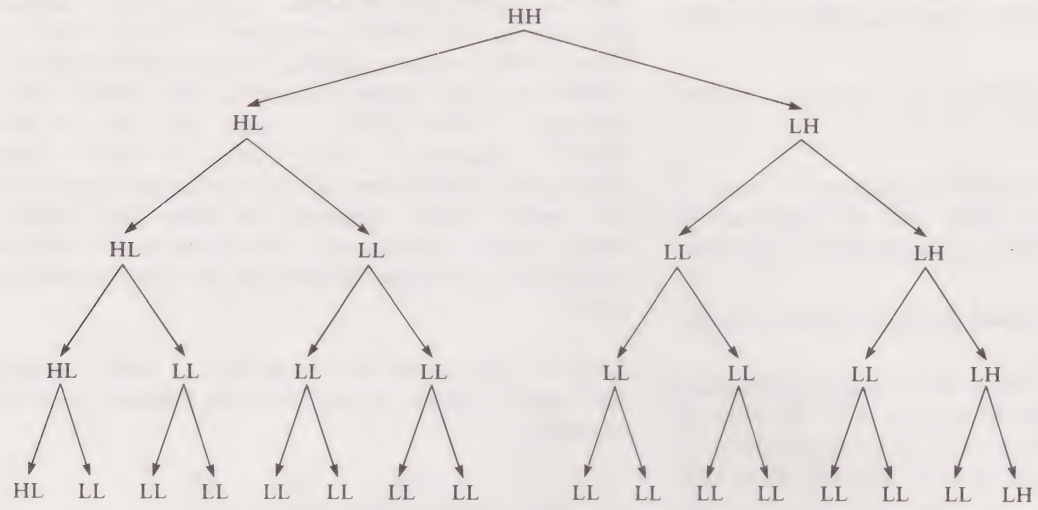



FIGURE 43 Base pairing in part of a DNA molecule during replication.

TABLE 9 For use with the answer to SAQ 8

| Semi-conservative replication | Generation | Ratio HL:LL |
|---|---|---|
|  | | |
| | 0 | – |
| | 1 | – |
| | 2 | 1:1 |
| | 3 | 1:3 |
| | 4 | 1:7 |



**SAQ 8** Parental generation 0 contained double helices with two heavy strands (HH). Semi-conservative replication resulted in generation 1 progeny all having DNA double helices with one heavy and one light strand (HL). Generations 2, 3 and 4 are formed as shown in Table 9.

As you see, two kinds of DNA exist in generation 2, HL and LL. These are in the ratio 1 : 1.

Generation 3 consists of HL and LL in the ratio 1 : 3.

Generation 4 consists of HL and LL in the ratio 1 : 7.

**SAQ 9** (a) Description (i) applies to both DNA and RNA.

(b) Descriptions (iii), (v) and (vi) apply to RNA but not to DNA.

(c) Descriptions (ii) and (iv) apply to DNA but not to RNA.

**SAQ 10** Table 10 is the completed form of Table 4.

**SAQ 11** (a) True. The ribosomes are the site of mRNA translation and the tRNA molecules bring the amino acids to the mRNA.

(b) False. Transcription is the process in which the genetic information of DNA molecules is transferred into mRNA molecules. Ribosomes are not involved in this.

(c) False. Amino acids do not bind directly to mRNA codons. Each amino acid attaches to its own tRNA. The specific tRNA (bearing its amino acid complexed to it) then binds to a specific mRNA codon.

(d) True. The site of mRNA synthesis is in the nucleus. mRNA is transcribed from the DNA in the nucleus but is translated in the cytoplasm.

(e) False. Each mRNA molecule can direct the synthesis of more than one molecule of the polypeptide for which it codes.

(f) False. Translation always starts at the 'left-hand end'

TABLE 10 For use with the answer to SAQ 10

| Type of nucleic acid | Where formed | Function(s) | Where function(s) takes place |
|---|---|---|---|
| DNA | inside the nucleus | encodes genetic information; passes it on to other cells by replication; passes it on to mRNA by transcription | inside the nucleus |
| mRNA | inside the nucleus | carries genetic information from nucleus to site of protein synthesis | in the cytoplasm |
| tRNA | inside the nucleus | forms complexes with specific amino acids, then binds to specific mRNA codons | in the cytoplasm |

of mRNA, and proteins are synthesized step by step from the N-terminal amino acid.

(g) True. Transcription of the genetic material has to occur in the nucleus as this is where the DNA is located in the cell.

(h) False. Genetic information is *transcribed* within nuclei.

SAQ 12 (a) 30 bases. As noted in Section 7.1, a set of three bases codes for one amino acid. So a sequence of ten amino acids is coded for by a sequence of 30 bases in mRNA.

(b) 30 bases. mRNA is formed by transcription from the coding strand of DNA.

(c) 18 kinds of tRNA. Every amino acid has (at least) one kind of tRNA molecule specific to it. Because 18 kinds of amino acid are involved in the synthesis of Q, at least 18 kinds of tRNA must be involved in its synthesis.

SAQ 13 Table 11 is the completed form of Table 7.

It is possible to fill in the first and last columns because both methionine (Met) and tryptophan (Trp) are each coded for by only one mRNA codon: AUG and UGG, respectively. This means that in all seven columns, one item is known within the first four rows; knowing one item, all the rest can be filled in by the rules of complementary base pairing.

SAQ 14 (a) 3; (b) 37; (c) 213.

(a) With six different kinds of base, a codon of one base would give six coding possibilities. If each codon contained two bases, there would be 36 ($=6^2$) possibilities, which is insufficient to code for the sequence of 37 amino acids. Therefore a codon must contain 3 (or more) bases, giving 216 ($=6^3$) coding possibilities.

(b) With 37 kinds of amino acid, the minimum number of kinds of tRNA molecule is 37.

(c) Because six kinds of base can provide 216 different triplet codons, and three of these codons are used as stop codons, 213 codons must code for amino acids. If every tRNA anticodon were different, then there would be 213 kinds of tRNA.

SAQ 15 If our hypothetical DNA contain an *uneven* number of bases, it would be impossible to conceive of a base-pairing system, because this requires bases to interact in pairs. SAQ 14 postulates that 'DNA replication and protein synthesis occur broadly in the ways described in Sections 5 and 6'. These ways depend on base pairing, so a system which has an uneven number of bases is not possible.

SAQ 16 The DNA sequence . . .AAA GAG. . . would give rise to the mRNA sequence . . .UUU CUC. . . . From Table 6 you can see that UUU codes for phenylalanine and CUC codes for leucine. The altered DNA sequence . . .TAA GAG. . . would give rise to the mRNA sequence . . .AUU CUC. . . , which—again referring to Table 6—would code for isoleucine (AUU) and leucine (CUC). Therefore, the amino acid phenylalanine would be missing at this point in the mutant protein and would be replaced by the amino acid isoleucine.

SAQ 17 The DNA sequences for the sixth, seventh and eighth codons would give the following mRNA sequence

| | 6th | 7th | 8th |
|---|---|---|---|
| DNA | CGA | TTC | AAA |
| mRNA | GCU | AAG | UUU |

which give the following sequence of amino acids in the normal protein:

alanine–lysine–phenylalanine

(a) Replacement of the first base of the sixth codon by G will create the DNA sequence GGA, and thus the CCU mRNA codon. This will now code for the amino acid proline. Although this alters the amino acid sequence of the enzyme, it is unlikely to have a drastic effect on the function of the enzyme.

(b) If the first base of the seventh codon is replaced with adenine, the DNA sequence will now be ATC and the mRNA codon will thus be UAG. This is a stop codon for the mRNA, and thus the polypeptide synthesis would stop at this point. The enzyme function would be zero as it could not be synthesized!

(c) Substitution of adenine by guanine in the DNA sequence will change the mRNA codon in the eighth position to UUC—this in fact makes no change to the amino acid incorporated. Thus you would expect no change in enzyme function.

(d) Deletion of two bases will cause misreading of all subsequent codons—thus a functional enzyme will not be produced.

(e) This will cause deletion of one amino acid. This will affect the structure of the enzyme, but as the deletion is distant from the active site the function of the enzyme may not be impaired.

TABLE 11 For use with the answer to SAQ 13

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ATG | AGG | TGT | TCG | CTC | GTA | TGG | non-coding strand } DNA double helix |
| TAC | TCC | ACA | AGC | GAG | CAT | ACC | coding strand } DNA double helix |
| AUG | AGG | UGU | UCG | CUC | GUA | UGG | transcribed mRNA |
| UAC | UCC | ACA | AGC | GAG | CAU | ACC | tRNA anticodon |
| Met | Arg | Cys | Ser | Leu | Val | Trp | amino acid incorporated into the polypeptide |

**SAQ 18** The seed colour of maize is determined by the pigments, and the pigments are produced through a series of biochemical interactions in which enzymes play an important part. In the white maize, a length of DNA codes for a different enzyme from that found in purple seeds. The presence of the different enzyme causes white rather than purple pigment to be produced.

**SAQ 19** (a) It is not always possible to determine the base sequence of the DNA, as the gene could have intron sequences. These sequences, although they are transcribed into mRNA, would be 'tailored out' before the mRNA was translated.

(b) As with (a), the gene could have intron sequences which would be present in the mRNA in the nucleus immediately after transcription. Thus the number of bases in the mRNA would be *more* than three times the number of amino acids. (Remember that each amino acid is coded by a codon of three bases.) While it is still in the nucleus, the mRNA will not yet have been cut and spliced (this happens in the cytoplasm).

(c) A large percentage of the DNA does not appear to have any coding function for RNA molecules. If there were a change in the DNA bases in this non-coding region of DNA, or even in the intron sequences within a gene, this would not be reflected in a change in the base sequence of any functional (or non-functional) mRNA produced.

(d) Rather than *changing* existing bacterial genes, the genes for human proteins are introduced into the bacteria and incorporated into their own genetic material as *additional* genes.

(e) Pre-pro-insulin has a pre-section which would be removed before storage. The central (C) section of the pro-insulin molecule has to be removed to release the A and B chains. These chains join up form the functional insulin molecule. The removal of both the pre-region and the central C section requires specific enzymes which would not be present in a cell-free system.

**SAQ 20** (i) Synthesize the DNA base sequences of the A and B polypeptide chains of insulin.

(ii) Incorporate these base sequences into bacterial plasmids, at a position adjacent to a specific bacterial gene.

(iii) Introduce these modified plasmids into *E. coli.*

(iv) Allow the bacteria to grow, and then make the bacteria transcribe the bacterial gene. Since this is linked to the insulin gene, a polypeptide is produced which has a bacterial protein and one of the insulin chains.

(v) Separate the bacterial protein from the insulin. Then mix the two insulin polypeptide chains (which have been 'grown' in different bacteria) together. Functional insulin—the joined A and B chains—is produced.

# ACKNOWLEDGEMENTS

Grateful acknowledgement is made to the following for permission to reproduce material in this Unit:

*Frontispiece* A. C. Barrington Brown.

*Figure 32* (*left*) From O. L. Miller *et al.* (1970) 'Visualization of bacterial genes in action' in *Science*, Vol. 169, © 1970 AAAS.

*Plates 1 and 2* courtesy of Eric Westhof, Centre Nationale de la Recherche Scientifique, IBMC, Strasbourg.

# INDEX FOR UNIT 24

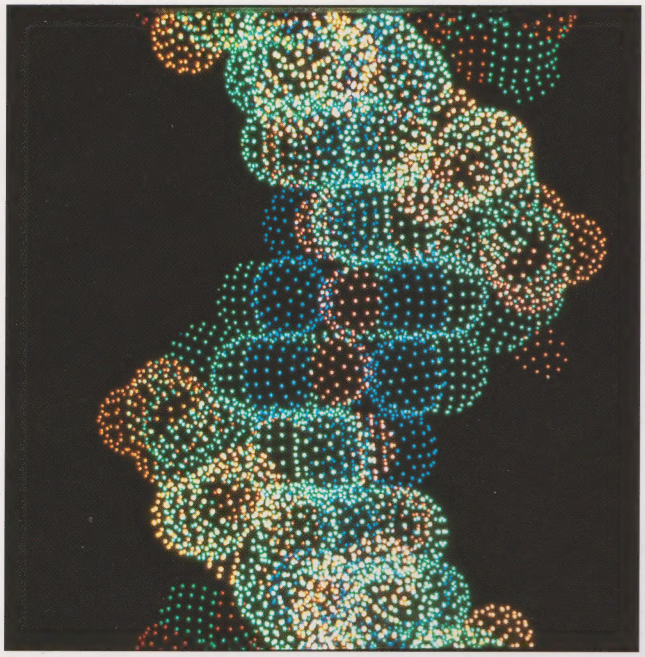


PLATE 1 Side view of DNA, with the individual atoms represented as dot-covered spheres. Carbon is green, oxygen is red, nitrogen is blue and phosphorus is yellow. The sugar–phosphate backbone on the far side of the structure is not shown.

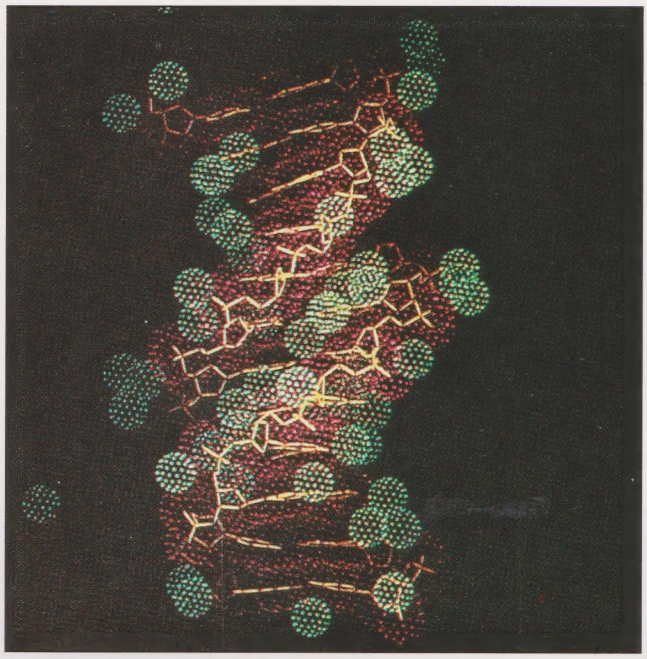


PLATE 2 Model of DNA. The skeleton is yellow with the surface indicated by red dots. The green dots show the location of the water molecules that would be found associated with the DNA in the nucleus.

PLATE 3a Bracken, *Pteridium aquilinum*, a fern that is common in Britain and widely distributed throughout the world.

PLATE 3b Common reed, *Phragmites australis*, a grass that grows in water and damp places and is common throughout the world.

PLATE 4 Roach, *Rutilus rutilus*, in an aquarium.